Contents

Review

Chinese Americans' Use of Patient Portal Systems: Scoping Review (e27924)
Katharine Lawrence, Stella Chong, Holly Krelle, Timothy Roberts, Lorna Thorpe, Chau Trinh-Shevrin, Stella Yi, Simona Kwon 3

Original Papers

Understanding the Potential of Mental Health Apps to Address Mental Health Needs of the Deaf and Hard of Hearing Community: Mixed Methods Study (e35641)
Judith Borghouts, Martha Neary, Kristina Palomares, Cinthia De Leon, Stephen Schueller, Margaret Schneider, Nicole Stadnick, Dana Mukamel, Dara Sorkin, Dakota Brown, Shannon Mc Cleerey-Hooper, Gloria Moriarty, Elizabeth Elkey 20

Process and Information Needs When Searching for and Selecting Apps for Smoking Cessation: Qualitative Study Using Contextual Inquiry (e32628)
Ylva Hendriks, Sebastiaan Peek, Maurits Kaptijn, Inge Bongers 31

Proposal for Post Hoc Quality Control in Instrumented Motion Analysis Using Markerless Motion Capture: Development and Usability Study (e26825)
Hanna Röhl, Patrik Atthoff, Radina Arsenova, Daniel Drebinger, Norman Gigengack, Anna Chorschew, Daniel Kroneberg, Maria Rönnefarth, Tobias Ellermeyer, Sina Rosenkranz, Christoph Heesen, Behnoush Behnia, Shigeki Hirano, Satoshi Kuwabara, Friedemann Paul, Alexander Brandt, Tanja Schmitz-Hübsch 54

Evaluation of the First Year(s) of Physicians Collaboration on an Interdisciplinary Electronic Consultation Platform in the Netherlands: Mixed Methods Observational Study (e33630)
Sanne Sanavro, Henk van der Worp, Danielle Jansen, Paul Koning, Marco Blanker, Prisma Platform Study Group 68

Implementation of a Web-Based Tool for Shared Decision-making in Lung Cancer Screening: Mixed Methods Quality Improvement Evaluation (e32399)
Julie Lowery, Angela Fagerlin, Angela Larkin, Renda Wiener, Sarah Skurla, Tanner Caverly 77

Requirements for a Bespoke Intensive Care Unit Dashboard in Response to the COVID-19 Pandemic: Semistructured Interview Study (e30523)
Brittany Davidson, Katiuska Ferrer Portillo, Marcell Wac, Chris McWilliams, Christopher Bourdeaux, Ian Craddock 90

A Participatory Design Approach to Develop Visualization of Wearable Actigraphy Data for Health Care Professionals: Case Study in Qatar (e25880)
Kamran Khowaja, Wafa Syed, Meghna Singh, Shahrad Taheri, Odette Chagoury, Dena Al-Thani, Michaël Aupetit 104
Smartphone Alcohol Use Disorder Recovery Apps: Cross-sectional Survey of Behavioral Intention to Use (e33493)
Rijuta Menon, Julien Meyer, Pria Nippak, Housne Begum. ................................................................. 129

Evaluating User Feedback for an Artificial Intelligence–Enabled, Cognitive Behavioral Therapy–Based Mental Health App (Wysa): Qualitative Thematic Analysis (e35668)
Tanya Malik, Adrian Ambrose, Chaitali Sinha. ................................................................. 138

Application of a Web-based Self-assessment Triage Tool During the COVID-19 Pandemic: Descriptive Study (e34134)
Anna Nowicka, Jakub Jaszczak, Anna Szymanek Pasternak, Krzysztof Simon. ................................................................. 148

Corrigenda and Addenda

Correction: Improving Pelvic Floor Muscle Training Adherence Among Pregnant Women: Validation Study (e38175)
Aida Jaffar, Sherina Mohd-Sidik, Chai Foo, Novia Admodisastro, Sobihatun Abdul Salam, Noor Ismail. ................................................................. 127
Review

Chinese Americans’ Use of Patient Portal Systems: Scoping Review

Katharine Lawrence1*, MD, MPH; Stella Chong2*, BA; Holly Krelle3*, MPhil; Timothy Roberts4*, MPH, MLS; Lorna Thorpe5, MPH, PhD; Chau Trinh-Shevrin2*, DrPH; Stella Yi2, MPH, PhD; Simona Kwon2, MPH, DrPH

1Healthcare Innovation Bridging Research, Informatics, and Design (HiBRID) Lab, Department of Population Health, New York University Grossman School of Medicine, New York, NY, United States
2Section for Health Equity, Department of Population Health, NYU Grossman School of Medicine, New York, NY, United States
3Division of Healthcare Delivery Services, Department of Population Health, NYU Grossman School of Medicine, New York, NY, United States
4NYU Health Sciences Library, NYU Grossman School of Medicine, New York, NY, United States
5Division of Epidemiology, Department of Population Health, NYU Grossman School of Medicine, New York, NY, United States
* these authors contributed equally

Corresponding Author:
Katharine Lawrence, MD, MPH
Healthcare Innovation Bridging Research, Informatics, and Design (HiBRID) Lab
Department of Population Health
New York University Grossman School of Medicine
227 E 30th St
6th Floor
New York, NY, 10016
United States
Phone: 1 6465013488
Email: katharine.lawrence@nyulangone.org

Abstract

Background: Electronic patient portals are increasingly used in health care systems as communication and information-sharing tools and show promise in addressing health care access, quality, and outcomes. However, limited research exists on portal use patterns and practices among diverse patient populations, resulting in the lack of culturally and contextually tailored portal systems for these patients.

Objective: This study aimed to summarize existing evidence on the access and use patterns, barriers, and facilitators of patient portals among Chinese Americans, who represent a growing patient population in the United States with unique health care and health technology needs.

Methods: The authors conducted a literature search using the PRISMA Protocol for Scoping Reviews (Preferred Reporting Items for Systematic Reviews and Meta-Analyses-ScR) for extracting articles published in major databases (MEDLINE, Embase, and PsycINFO) on patient portals and Chinese Americans. Authors independently reviewed the papers during initial screening and full-text review. The studies were analyzed and coded for the study method type, sample population, and main outcomes of interest.

Results: In total, 17 articles were selected for inclusion in the review. The included articles were heterogenous and varied in their study aims, methodologies, sample populations, and outcomes. Major findings identified from the articles include variable patterns of portal access and use among Chinese Americans compared to other racial or ethnic groups, with limited evidence on the specific barriers and facilitators for this group; a preference for cross-sectional quantitative tools such as patient surveys and electronic health record–based data over qualitative or other methodologies; and a pattern of aggregating Chinese American–related data into a larger Asian or Asian American designation.

Conclusions: There is limited research evaluating the use patterns, experiences, and needs of Chinese Americans who access and use patient portal systems. Existing research is heterogeneous, largely cross-sectional, and does not disaggregate Chinese Americans from larger Asian demographics. Future research should be devoted to the specific portal use patterns, preferences, and needs of Chinese Americans to help ensure contextually appropriate and acceptable design and implementation of these digital health tools.
patient portal; electronic health records; personal health records; eHealth; health equity; digital divide; Chinese Americans; Asian Americans

Introduction

The expansion of health information technology (HIT) has provided patients with tools to proactively access their health information, self-manage chronic conditions, and communicate directly with providers [1]. In particular, electronic patient portals— which are secure internet-based platforms or websites that provide patients with 24-hour access to their personal health information—have emerged as a common communication and information-sharing tool for health care systems [2]. Patient portals offer a variety of features and functions for patients, such as the ability to access and review medical information, view lab and imaging results, schedule medical appointments and other visits, and interact with their health care providers [2-4]. Increasingly, these systems are directly integrated into electronic health record (EHR)-based platforms (eg, Epic MyChart or eClinical Works) or customer relationship management systems, as well as into the growing ecosystem of telehealth services. The COVID-19 pandemic expanded the use of patient portals as a facilitator of virtual health care and telemedicine, remote patient-provider communication, and monitoring [5-7]. Patient portals have demonstrated effectiveness in improving patient communication, engagement, and satisfaction [8,9], with some evidence on improvements in health outcomes [7,10] and lowered health care costs [6]. However, despite these benefits, adoption of and engagement with patient portals have varied, and significant disparities in the use of portal systems have been identified [2,11-18]. These disparities are shaped by individual, community, and structural factors such as social demographics (eg, socioeconomic status), health status (eg, disability diagnosis, chronic illness status), human-computer interface design challenges (eg, usability), and structural barriers (eg, lack of access to broadband internet).

Chinese Americans are a population frequently under- or mis-represented in health care, health delivery, and health research [19,20]. At roughly 5 million people, Chinese Americans comprise the largest subgroup of a heterogeneous community of Asian Americans and Pacific Islanders (AAPI), who themselves represent almost 10% of the US population [21-23]. Chinese American patients have distinct experiences interacting with the health care system [23,24], including care moderated by health technologies [25-27]. Although health disparities in this community have been identified and are mediated by factors such as language proficiency and immigration status [22,24], the details of these experiences are often obscured by problems with data collection and interpretation of health data that ignores the considerable heterogeneity and complexity of the AAPI designation [28].

To improve the effectiveness, acceptability, and use of digital health technologies such as patient portals among diverse communities, a better understanding of the use patterns and practices of the specific communities and their subgroups is needed. This scoping review summarizes the existing evidence on patient portal perceptions, adoption, and use among Chinese Americans, and it highlights gaps and areas for further research on patient portal and digital health technology use among Chinese Americans and other diverse patient populations.

Methods

The aim of conducting a scoping review is to identify and broadly describe knowledge and research pertaining to a topic of interest as well as to identify trends, patterns, and gaps in the literature. Scoping reviews are ideal for research areas where the study question is broad or exploratory, there is limited literature on the topic, or study methodologies are diverse [29].

The review was conducted following the PRISMA Protocol for Scoping Reviews ([Preferred Reporting Items for Systematic Reviews and Meta-Analyses-ScR] [30]). In August 2020 and 2021, one of the coauthors (TR) who is an experienced medical librarian searched MEDLINE, Embase, and PsycINFO using the Ovid Platform and the Web of Science Core Collection. The search was not limited by language or publication date. Quantitative and qualitative studies that included primary data collection or data analysis were included; article types such as opinion pieces or letters to editors were excluded. The complete Ovid MEDLINE search strategy is available in Multimedia Appendix 1.

US-based studies that described the inclusion and perceptions of Asian Americans (eg, Asians, Asian Americans, Chinese Americans, and Filipinos) toward electronic patient portals were included. Studies that identified Asian Americans only under the heading of “Other” without additional specificity were excluded. Patient portals were defined as web-based platforms that provided access to data from EHRs, including features such as medical histories, visit summaries, medication lists, as well as secure messaging features, access to educational resources, and appointment scheduling [19,20,31]. Studies that focused primarily on the delivery of “real-time interactive” remote clinical care using audio or video communication technology (eg, synchronous telemedicine) [32] were excluded, as these technologies often exist separately from patient portal communication systems or do not support key asynchronous features such as personal health data review by patients or remote monitoring. Studies exploring general health information literacy or information-seeking via digital resources (eg, the internet) in this group were also excluded.

After duplicates were removed, 1505 articles remained. Titles and abstracts were screened using Covidence software [33] by 2 independent reviewers (SKC and HK) for explicit or implicit mention or identification of Chinese Americans. Conflicts were resolved through discussion between the 2 reviewers until consensus was reached. When needed, consultation was sought from another coauthor (KL) to reach consensus. The full texts,
including tables, figures, and appendices, of 65 articles were reviewed following the same process. Ultimately, 17 articles were included in the analysis, as shown in Figure 1.

**Figure 1.** PRISMA Flowchart showing the screening and inclusion process of the studies. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

**Results**

**Article Summaries**

In total, 17 articles were selected for inclusion in the review. A summary of each article, including the study design, sample information (including the level of Asian American population identified in the study), and key findings can be found in Tables 1 and 2.
Table 1. Characteristics of studies that include identifiable data specific to Chinese Americans.

<table>
<thead>
<tr>
<th>Study</th>
<th>Objective</th>
<th>Research design/tools</th>
<th>Sample population/level of Chinese American granularity and location</th>
<th>Patient portal technology/feature</th>
<th>Relevant results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ackerman et al (2017) [33]</td>
<td>To understand the implementation of patient portals in safety net health care systems striving to meet MU criteria set by the Federal United States government</td>
<td>Mixed methods Rapid ethnography to assess MU, including interviews with providers and executives, informal focus groups with frontline staff, observations of patient portal sign-up procedures, and review of marketing materials and patient portal use. Administered modified version of the American Medical Association’s Health IT Readiness Survey Study tools: patient portal promotional flyers in English and Chinese at clinics; instructional video in Cantonese; language-congruent health staff available</td>
<td>5 California safety net health systems. Site 5 in Northern California, which serves 95% non-native English-speaking Chinese immigrants. Location: Northern California</td>
<td>Patient portal (NextGen) implementation strategies and efforts at Site 5 Patient portal features: medical history, test results, secure messaging, and appointment requests</td>
<td>Overall: Community health centers were motivated by MU incentives to increase patient portal enrollment and integrate portal-related work into clinic routines. Barriers to patient portal usage for patients: lack of internet access, lack of computer proficiency, discomfort with portal use, language barriers, fear of government surveillance, and preference for in-person interaction with providers. Specific to Site 5: Chinese American patients face language barriers in accessing the patient portal. “The (EHR vendor) website isn’t in their Chinese language… How were they going to get their patients to be able to utilize this?” Perception that clinic discouraged staff from promoting patient portal once MU threshold was reached.</td>
</tr>
<tr>
<td>Study</td>
<td>Objective</td>
<td>Research design/tools</td>
<td>Sample population/level of Chinese American granularity and location</td>
<td>Patient portal technology/feature</td>
<td>Relevant results</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------</td>
<td>-----------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Gordon and Hornbrook (2016)   | To identify racial or ethnic and age disparities among older patients’ use of patient portals and access to digital technology and devices for email and web-based health care management programs | Quantitative (cross-sectional, administrative data and survey)                        | Study 1: Analyzed administrative data about patient portal account status and use from the KPNC health plan  
Study 2: Mailed English survey questionnaire, from 2013 to 2014, to stratified random sample of Study 1’s population | KPNC internet-based patient portal, kp.org, and other digital health technology and tools (eg, emails, text, computer, smartphones) | Compared to Black, Filipino, and Latino older patients, Chinese and non-Hispanic White older patients were more likely to be registered to use the patient portal and more likely to use portal functions.  
Chinese and non-Hispanic White older patients have higher levels of access to digital tools, experience in performing a variety of web-based tasks, and belief in their ability to seek health information on the internet compared to Black, Latino, and Filipino peers.  
Chinese older people prefer having telephone appointments with health coaches and are less interested in reading about health topics on the internet.  
Chinese older people have the lowest level of interest in using health apps. |
| Gordon and Hornbrook (2018)   | To assess disparities by race/ethnicity and age on older patients’ ability to engage with online health information and mobile health tools connected to their health system | Quantitative (cross-sectional, survey)                                               | Stratified random sample of 5420 English-speaking KPNC patients  
Chinese (n=500), non-Hispanic White (n=1420), African American/Black (n=1500), Hispanic/Latino (1500), and Filipino (n=500)  
Location: Northern California | Digital health technology and tools (eg, internet, computer, mobile phone, email, text, social media, apps) |                                                                                                                                                                                                                   |
<table>
<thead>
<tr>
<th>Study</th>
<th>Objective</th>
<th>Research design/tools</th>
<th>Sample population/level of Chinese American granularity and location</th>
<th>Patient portal technology/feature</th>
<th>Relevant results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Khoong et al (2020) [36]</td>
<td>To assess predictors of health technology use (eg, language preferences, smartphone ownership, type of clinic for health care)</td>
<td>Quantitative (cross-sectional, survey)</td>
<td>Nonrandom sample of 1027 participants</td>
<td>Digital health technology and tools for communication with clinicians (eg, email, text, phone apps, web-based health videos, and online health support groups)</td>
<td>Relative to English-speaking survey respondents, individuals who preferred the Chinese language had lower odds of texting or using an app to communicate with their clinician. There were no differences in using emails or watching web-based health videos. Language concordance was suggested as a major barrier.</td>
</tr>
</tbody>
</table>

aMU: meaningful use.
bIT: information technology.
cEHR: electronic health record.
dKPNC: Kaiser Permanente Northern California
<table>
<thead>
<tr>
<th>Study</th>
<th>Objective</th>
<th>Research design</th>
<th>Sample population/ level of Chinese American granularity and location</th>
<th>Focus</th>
<th>Relevant results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ahlers-Schmidt and Nguyen (2013) [37]</td>
<td>To obtain parents’ feedback and intention to use patient portals for their children’s health records and concerns post the facilitated learning session</td>
<td>Quantitative (cross-sectional, survey)</td>
<td>Parents of patients. (N=65) White (n=26, 40%); Hispanic (n=14, 22%); Asian (n=9, 14%); African American (n=6, 9%); Mixed/other race (n=8, 12%) Location: Kansas</td>
<td>University of Kansas Pediatric Clinic’s eClinical Works, an electronic medical record with a patient portal</td>
<td>Most parents did not know about the patient portal before the study demonstration. Parents expressed that patient portal was simple to use after demonstration. Parents liked portal functions such as viewing lab results and medical records; disliked need to make separate accounts for each child and the lack of a symptom checker function.</td>
</tr>
<tr>
<td>Dalrymple et al (2018) [38]</td>
<td>To assess parents’ use of the internet for health information and parents’ awareness of digital health technologies to obtain health information Screening questions assess parents’ level of health literacy and interest in use of patient portals</td>
<td>Quantitative (cross-sectional, survey) Study tool: 26-question paper and pencil survey adapted from interview protocol designed from previous study</td>
<td>Total sample population of parents or adult caregivers of children and adolescents, N=270 Asian (1.9%); American Indian/Alaska Native (1.5%); Black/African American (38.1%); Hispanic/Latino (13.7%); Native Hawaiian/Pacific Islander (0.4%); White (40.7%); more than one race/ethnicity (4.4%); and Other (1.5%) Location: Unspecified large metropolitan area in eastern United States</td>
<td>Internet and patient portal</td>
<td>Most patients reported having access to the internet and using the internet to seek general and health information. Respondents expressed enthusiasm and interest in using a patient portal if it were available from their health care provider.</td>
</tr>
<tr>
<td>Foster and Krasowski (2019) [39]</td>
<td>To assess patient portal usage by ED patients at an academic medical center using patient portal activation rates and rates of accessing diagnostic test results on patient portals</td>
<td>Quantitative (retrospective cohort, EHR, and administrative data)</td>
<td>25,361 unique ED patients identified via EHR patient portal records Asian (n=451); African American/Black (n=2,254); White (n=20,637); Hispanic/Latino (n=1,257); Other (n=762) Location: Iowa</td>
<td>UIHC’s patient portal (MyChart), connected to EPIC EHR system</td>
<td>Highest rates of using the patient portal to view laboratory and radiology results were observed for younger female, proxies, Asian, and White patients. Activation rates were highest for Asian and White patients. Disparities were observed among teenagers, older adults, African American/Black, and Hispanic/Latino patients.</td>
</tr>
<tr>
<td>Study</td>
<td>Objective</td>
<td>Research design</td>
<td>Sample population/ level of Chinese American granularity and location</td>
<td>Focus</td>
<td>Relevant results</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Goel et al (2011) [12]</td>
<td>To examine the enrollment in and use of patient portal at an academic medical center by race/ethnicity, gender, and age</td>
<td>Quantitative (cross-sectional, EHR and administrative data)</td>
<td>Patients enrolled in the patient portal system, N=7088</td>
<td>Northwestern Medical Faculty Foundation’s EHR patient portal</td>
<td>Significant disparities in patient portal enrollment by race/ethnicity were observed, but not by age or gender. White patients (74%) were more likely to enroll in patient portals compared to Black (55%), Latino (64%), and Asian (66%) patients. When adjusted for variables (eg, age, gender, income, education, and provider effects), the disparity between Asian and White patients was no longer statistically significant.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Study tool: patients’ use of EHR-based advice function and request for refills</td>
<td>Asian (n=142, 2%); White (n=3472, 49%); Black (n=1063, 15%); Latino (n=284, 4%); Other (n=851, 12%); Missing race/ethnicity (n=1347, 19%)</td>
<td>Location: Chicago</td>
<td></td>
</tr>
<tr>
<td>Graetz et al (2016) [40]</td>
<td>To assess sociodemographic disparities in patient portal use</td>
<td>Quantitative (cross-sectional, survey)</td>
<td>Total study participants from KP-NC, N=1041</td>
<td>Internet and email</td>
<td>Asian and Black respondents were more likely to rarely or never to use the internet (45.4% and 45.6%, respectively) compared to their White respondents. Asian participants (78%) preferred in-person care over telephone care compared to White patients (64%).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Study tool: Administered paper-based survey mailed to participants; survey measures internet access, secure email use, care preference, sociodemographics, and health characteristics</td>
<td>White (n=617, 59.3%); Asian (n=145, 13.9%); Black (n=122, 11.7%); and Hispanic (n=128, 12.3%)</td>
<td>Location: Northern California</td>
<td></td>
</tr>
<tr>
<td>Ketterer et al (2013) [41]</td>
<td>To identify predictors of patient portal enrollment and activation among a pediatric primary care population</td>
<td>Quantitative (cross-sectional, EHR and administrative data)</td>
<td>Total sample population N=84,015</td>
<td>Patient portal site, MyNemours</td>
<td>Adjusted odds of portal enrollment were lower for Asian respondents compared to White respondents. Once enrolled, there was no difference in portal activation between Asian respondents and White respondents. Study suggested language concordance as a major barrier.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Study tool: primary care database, and enrollment in and use of a patient portal</td>
<td>Black (n=35,286, 42%); Asian (n=35,286, 42%); White (n=35,2520, 3%); Hispanic (n=10,082, 12%); Other (n=9,242, 11%); and Unknown (n=1,680, 2%)</td>
<td>Location: Delaware</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Objective</td>
<td>Research design</td>
<td>Sample population/ level of Chinese American granularity and location</td>
<td>Focus</td>
<td>Relevant results</td>
</tr>
<tr>
<td>----------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>------------------------------------------------------</td>
<td>-----------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Lyles et al (2013) [42]</td>
<td>To understand how patient-provider relationships influence patients’ use of online patient portals and secure messaging</td>
<td>Quantitative (cross-sectional, survey)</td>
<td>Surveyed patients DISTANCE® Black (23%); Latino (16%); East Asian (ie, Chinese, Japanese, Korean, or Vietnamese) (10%); Filipino (12%); and Other (6%)</td>
<td>KPNC’s internet-based patient portal, kp.org</td>
<td>White and Latino individuals with higher trust in the providers were more likely to register on the patient portal. There was no relationship between trust in provider and patient portal use for Asian respondents.</td>
</tr>
<tr>
<td>Lyles et al (2016) [43]</td>
<td>To determine whether racial/ethnic minority patients’ use of the patient portal’s medication refill function has changed over time compared to White patients</td>
<td>Quantitative (EHR and administrative data) Study tool: diabetic patients’ use of EHR-based medication refill function</td>
<td>White (58%); Asian (10%); Latino (9%); Filipino (9%); Black (7%); and Mixed/other (9%)</td>
<td>KPNC’s internet-based patient portal, kp.org</td>
<td>Asian were not less likely to exclusively use refill functions than other ethnic groups. Adherence to medication refills improved over time for all ethnic groups, but there was no significant difference between ethnicities. Usage and accessibility were identified as barriers to portal registration.</td>
</tr>
<tr>
<td>Miles et al (2016) [44]</td>
<td>To measure and evaluate the frequency at which patients use the patient portal to view online radiology reports</td>
<td>Quantitative (cross-sectional, EHR and administrative data) Study tool: patient interactions with portal features (eg, radiology, laboratory, and clinical notes) and sociodemographic factors</td>
<td>Asian or Pacific Islander (n=6376, 10.4%); American Indian or Alaska Native (n=522, 0.8%); Black or African American (n=3817, 6.2%); Hispanic or Latino (n=1850, 3%); White (n=44,163, 72.25); and Other/more than one race (n=675, 1.1%); and Unknown (n=3728, 6.1%)</td>
<td>UW’s® patient portal system, UW eCare web portal</td>
<td>Asian respondents were more likely than White patients to view their radiology reports. Older patients, primary non-English speakers, and those with non-commercial insurance viewed reports at lower rates. Concerns identified in the study include loss of patient confidentiality, health information inaccuracy, and disruption of patient-physician relationship.</td>
</tr>
<tr>
<td>Patel et al (2011) [45]</td>
<td>To determine low-income, ethnically diverse consumers’ attitudes and beliefs toward HIE® and use of HIE via PHRs® and to identify factors that impact consumers’ support for providers’ use of HIE and their own personal use of PHRs</td>
<td>Quantitative (cross-sectional, survey) Study tool: survey adapted from previously validated national surveys. Survey was translated into Spanish, Russian, and Mandarin Chinese</td>
<td>BHIX®’s patients White (n=36, 74%); Asian (n=57, 28%); African American (n=20, 10%); and Other (n=56, 27%). Spoke Chinese at home (n=42, 20%)</td>
<td>EHRs, internet, HIE, and PHRs</td>
<td>Compared to other racial/ethnic groups in the study, Asian Americans indicated lower levels of support for HIE (48%) and lower levels of potential PHR usage (67%).</td>
</tr>
<tr>
<td>Study</td>
<td>Objective</td>
<td>Research design</td>
<td>Sample population/ level of Chinese American granularity and location</td>
<td>Focus</td>
<td>Relevant results</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>-------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Sarkar et al (2010) [46]</td>
<td>To examine whether use of an internet-based patient portal differed between English-speaking patients with limited health literacy and English-speaking patients with adequate health literacy</td>
<td>Quantitative (cross-sectional, survey) Study tool: DISTANCE study was conducted in English, Spanish, Cantonese, Mandarin, and Tagalog</td>
<td>Total of 14,201 surveyed participants from DISTANCE study Non-Hispanic White (n=3957, 28%); Latino (n=1923, 14%); African American (n=2899, 21%); Asian (n=1253, 9%); Filipino (n=1624, 12%); Other (n=2446, 17%) Location: Northern California</td>
<td>KPNC’s internet-based patient portal, kp.org</td>
<td>Study did not find increased risk of not signing onto the patient portal for Asian Americans compared to African American, Latino, and Filipino respondents. Asian Americans had lower rates of never using patient portal functions including lab result viewing, medication refills, email, and scheduling appointments. Health literacy was identified as a barrier to portal activity.</td>
</tr>
<tr>
<td>Sarkar et al (2011) [47]</td>
<td>To examine portal use habits via the frequency at which participants requested a password for the patient portal, the proportion of participants who activated their accounts by changing the default password, and the proportion of participants who login to their accounts using their personal, customized password</td>
<td>Quantitative (cross-sectional, EHR and administrative data) DISTANCE study was conducted in English, Spanish, Cantonese, Mandarin, and Tagalog</td>
<td>Total of 14,201 surveyed participants from DISTANCE Study Non-Hispanic White (n=3957, 28%); Latino (n=1923, 14%); African American (n=2899, 21%); Asian (n=1253, 9%); Filipino (n=1624, 12%); Other (n=2446, 17%) Location: Northern California</td>
<td>KPNC’s internet-based patient portal, kp.org</td>
<td>Asian American (53%) and White (51%) participants were more likely than their African American (31%), Latino (34%), and Filipino (32) counterparts to request a password for the internet-based patient portal and to login to the patient portal after requesting a password. Older adults with less educational attainment were less likely to register and use the patient portal.</td>
</tr>
</tbody>
</table>
Relevant results

Sample population/ level of Chinese American granularity and location

Focus

Study | Objective | Research design | Sample population/ level of Chinese American granularity and location | Relevant results
--- | --- | --- | --- | ---
Tieu et al (2017) [48] | To measure participants’ satisfaction with use of patient portal | Mixed methods (cross-sectional, usability testing and survey) | Total of 25 English-speaking (23 patients and 2 caregivers) participants. African American (n=9, 36%); White (n=6, 24%); Hispanic (n=2, 8%); Asian or Pacific Islander (n=5, 20%); and Other (n=3, 12%). Location: San Francisco, California | RFPC’s patient portal, MYSFHEALTH

Participants with limited health literacy, including Asian and Pacific Islander patients were more likely to need assistance navigating the patient portal. Barriers to patient portal use for participants with limited health literacy include (1) lack of basic computer skills; (2) routine computer use challenges despite basic knowledge of computers; (3) difficulty reading, writing, and understanding language; and (4) difficulty understanding and applying medical information from the internet and patient portal.

The included articles varied in terms of the study methodology, sample population, data collection methodology, and geographic area within the United States. Among these, 10 were from populations in California [34-36,40,42,43,45-48], with 5 from the Kaiser Permanente health system [35,36,42,43,46]. Further, 3 studies used a shared database—the Diabetes Study of Northern California (DISTANCE)—to analyze portal-related outcomes [40,47,48]. Of the data collection tools described in these studies, 5 studies indicated they were available and conducted in Chinese (eg, Mandarin or Cantonese) [12,34,43,47,48].

Overall, the articles described heterogenous results among varied patient populations, health conditions, and care settings. Few clear themes emerged and results specific to Asian American subgroups such as Chinese Americans were not identified. In general, the authors were able to identify the following major themes and trends from the results.

1. Chinese Americans demonstrate variable patterns of patient portal access and use as compared to other demographics, particularly racial or ethnic groups; exploration of the specific contexts of use, including barriers and facilitators, is limited.
2. Most studies employed cross-sectional, quantitative tools to assess patient portal use patterns and practices, including patient surveys and EHR-based data that measure portal activity (eg, logins and click-throughs); neither longitudinal nor significant qualitative research studies were conducted to validate or further explore nuances in findings specific to Chinese Americans.
3. Despite the heterogeneity of the populations included in AAPI designation, studies exploring patient portals do not disaggregate Asian and Asian American study populations into Chinese Americans and other subgroups.

Findings Specific to Chinese Americans

Only 4 studies [33-36] specifically disaggregated Chinese American populations (Table 1). All 4 of these were from California. Among these, 3 [34-36] were primarily based around surveys, and 1 [33] was based on rapid ethnography, mostly focusing on understanding the barriers to accessing patient...
ports. Barriers reported included language barriers, lack of internet access or computer proficiency, fear of government surveillance, and a preference for in-person interaction. Further, 2 of the studies [34,35] found that Chinese patients were more likely than other non-White groups to register and use internet-based portals, and 1 [40] found that relative to English-speaking respondents, people who preferred the Chinese language were less likely to send text messages or use an app to contact their clinician.

**Chinese Americans Demonstrate Variable Patterns of Patient Portal Access and Use Compared to Other Racial or Ethnic Groups**

This represents a finding in the data across studies, with some demonstrating lower rates of use and others demonstrating higher rates and rates comparable to White patients. In a study on the use of the Northwestern Medical Faculty Foundation’s electronic patient portal [44], the authors found that once variables such as age, gender, education, income, and provider effects were adjusted, there was no disparity between the enrollments of Asian American and White patients on the patient portal. In another study of Chinese American older adults in Kaiser Permanente, Northern California [35], the authors found that non-Hispanic White and Chinese American older adults were more likely than other racial or ethnic groups to register for using the portal and its functions such as sending messages, viewing lab results, or ordering prescription refills. Other studies showed lower use and lower motivation to use digital health technology among Chinese Americans. In their study examining patients’ patterns of texting and communication with their clinicians via apps, Khoong et al [36] found that individuals who preferred to use Chinese language had lower odds of texting or using an app to communicate with their clinicians compared to English-speaking survey respondents. In a study assessing older patients’ readiness to use eHealth tools, researchers found that Chinese American patients had the lowest level of interest in using patient portal technology among all the racial or ethnic groups in the study, though their experience of using the internet was similar to that of non-Hispanic White patients [36]. In their assessment of attitudes toward health information exchanges (HIEs) and personal health records (PHRs), Patel et al [45] found that Asian Americans were less likely than other racial or ethnic groups to support the use of PHR technology.

Identified studies provided limited evidence on the barriers faced by Chinese Americans in using patient portals. For individuals, the main reported barrier was language congruency with the portal or related technologies, or English language proficiency. In a mixed methods study evaluating the implementation of meaningful use at community health centers in California, Ackerman et al [33] noted that many patients could not read English and that even if communication with care providers could be conducted in Chinese, most EHR features (including records, test results, and communication tools like the patient portal) were exclusively in English. The authors also noted concerns among some Chinese Americans about government surveillance, particularly among patients who were undocumented or had concerns regarding their immigration status. Additional individual-level barriers identified in the studies included issues of usability and accessibility of the portal tool [43], concerns around confidentiality and privacy [38], low health literacy, [48], and digital literacy [45]. Conversely, in a study assessing the influence of patient-provider relationships on patient portal and messaging usage, Lyles et al [42] found that although trust in providers was correlated with registration for portals by White and Latinx patients, this was not the case for Asian patients.

Identified community and structural barriers were largely related to clinic-level resources and included the clinical staff’s ability to support patients’ engagement in patient portal technology and the paucity of language-congruent support services. In their rapid ethnography with clinical staff in safety net hospital–affiliated practices, Ackerman et al [33] reported challenges related to providers and staff members having limited time and skills to coach patients in using the patient portal, and concerns regarding meaningful use metrics that prioritize outcomes such as portal sign-up rather than sustained use. The researchers also identified disruptions to clinical workflows and increased administrative burden as barriers to effective implementation and use of EHR-related tools. In 3 studies, access to digital technology and infrastructure such as the internet was associated with higher rates of patient portal access and use by Chinese and Asian American patients [35-37].

**Most Studies Employed Cross-sectional Quantitative Tools to Assess Patient Portal Use Patterns and Practices, Including Patient Surveys and EHR-Based Data That Measure Portal Activity**

Among the 17 studies, 8 employed survey-based, numeric (eg, Likert scale) data collection tools disseminated using either digital tools (eg, email) or in person. Survey question areas ranged from portal familiarity and general perspectives to personal experiences, feature preferences, and self-reporting of details on use habits [12,55–37,39,40,43,47]. The remaining studies used either administrative information–based EHRs or associated databases. Furthermore, 6 studies conducted primary EHR-based analyses to identify patterns and trends in portal-based activities [38,41,44,46,48,49]. Key EHR- and portal-based measures reported by researchers included patient portal registrations [35,44], logins and appointment booking [47–49], medication refill requests [46], viewing of results and reports (eg, radiology reports) [38,41,47], and texting and other forms of communication with clinicians [40,42]. These activities were analyzed for frequency and other patterns, and they were often compared among demographics such as age, race or ethnicity, sex or gender, income level, insurance status, and language. Key themes in the survey questions included actual and expected use of different features, concerns and barriers related to using portals, and confidence in the ability to use portals and understand health information shared through these portals. Most of these measures are applied cross-sectionally, and there is neither longitudinal nor significant qualitative research to validate or further explore nuances in findings specific to Chinese Americans or other Asian American subgroups. No studies included measures of associated health outcomes.
Despite the Heterogeneity of the Populations Included in AAPI Designation, Studies Exploring Patient Portals Largely do not Disaggregate Asian and Asian American Study Populations

Of the 17 studies included in this review, only 4 specifically disaggregate or discuss Chinese Americans [34-36,40]. The remaining studies generally refer to “Asian Americans” or “Asians,” with only indirect references to over 20 unique ethnic subgroups included in that designation or otherwise included in the study sample, data collection, or analysis. For example, Chinese-speaking patients were occasionally mentioned in the text or tables of these studies [12,34,43,47,48] but not included in any multivariate analyses as a separate category. In these studies, it was inferred that Chinese American patients were included via references to the languages of the data collection instruments (eg, Mandarin or Cantonese) or the study database being used for analysis. No studies specifically or exclusively evaluated Chinese Americans’ attitudes toward, perceptions about, or use of patient portal technology.

Discussion

Principal Findings

This scoping review highlights the extremely limited research on the use patterns, experiences, and needs of Chinese Americans who access and use patient portal systems for their health care. The identified studies were heterogenous in their approaches and outcomes, making generalizable trends in the data difficult to identify, although we were able to identify some patterns in the research methodologies and data collection tools across studies. By and large, the existing studies have focused on the identification of varying portal use patterns among racial, ethnic, and other demographics, and their correlative predictors such as age, primary language, or health literacy. Overall, the studies obtained mixed findings regarding the rates of portal usage by Chinese Americans when compared to other populations, with some indicating lower rates of portal adoption and use when compared to White patients and others finding comparable rates. We were unable to identify trends more granularly in terms of portal access within Chinese American subgroups (eg, women, geographic populations) due to limitations in the available data. We identified individual- and system-level factors that contributed to use patterns, as well as barriers to access and usage. Relevant individual-level factors included English language proficiency and language congruency with portal technology; health literacy; perceived usability and usefulness of the technology; and trust in provider relationships, privacy, and confidentiality. Relevant system-level factors included clinical resource and capacity limitations, and access to digital tools such as email and the internet. Studies tended to be cross-sectional and quantitative in nature, with minimal exploration of longitudinal trends in use patterns or practices, qualitative aspects, or correlation with health outcomes. Finally, we identified a pattern of data aggregation practices that tended to combine and compare Asian Americans as a larger demographic group to other racial or ethnic groups, rather than identifying data at the level of Chinese Americans or other subgroups. This practice had the effect of generalizing learning across Asian Americans, thus providing limited insight into the experiences of Asian subgroups of different ethnicities, languages, and religious affiliations, among other factors.

To our knowledge, this is the first study to evaluate the patient portal use patterns and needs of Chinese Americans. Prior research has explored various features of patient portal activity, use, and experience in other clinical contexts, including among Black and Latinx communities and vulnerable populations such as the older people and those with disabilities [50-52]. A comprehensive review of interventions to increase patient portal use in “vulnerable populations” by Grossman et al in 2019 [4] identified 18 studies evaluating the impact of interventions designed to increase portal use or reduce disparities in use. The authors noted that most studies focused on individual-level interventions such as patient education and training and identified a lack of interventions or programs targeting tool-use (eg, patient portal interfaces or features), community-, organizational-, or system-level factors to improve portal adoption and use [4]. This is also supported by the findings of the study led by Antonio et al [52] that explored patient portal research through the lens of health equity and identified a varying and often superficial level of interest in portal technology among underserved groups by researchers and an underemphasis on the systemic factors influencing patient portal access and use among diverse communities. Although comprehensive, these reviews included limited information on the needs, use patterns, or potential interventions for specific vulnerable groups, particularly among racial or ethnic demographics; as observed in our findings, data on race and ethnicity included in these reviews often excluded Asian Americans or did not identify Asian American subgroups. Though our study includes some of the articles referenced by these reviews, our focus on Chinese and Asian American subgroups provides additional specificity to the overall literature on patient portals and exposes existing challenges in identifying and applying appropriately tailored solutions to technical problems for undifferentiated “vulnerable” patients.

The findings of this study have important implications for the design and deployment of patient portals and other digital health tools (eg, EHRs, mobile health apps) as well as for the study of health technology usage among Chinese Americans, Asian subgroups, and other diverse or vulnerable patient populations. Overall, there is need for a more granular study focusing on the use of digital health technology by diverse communities to elucidate key differences in their needs, preferences, and constraints. Participatory design frameworks that incorporate diverse stakeholders to identify and address specific needs, preferences, and concerns regarding health care technologies can help inform more effective and sustainable implementation of these tools in clinical practice. Frameworks and methodologies that explicitly address digital health disparities and digital health equity, such as the equity-centered design framework [53] and the digital health equity framework [54], can additionally help identify and overcome structural barriers such as access to digital infrastructure or institutional racism. At the same time, there is a need for clearer definitions and more granular breakdowns of populations included in data collection and data publication processes to better inform
appropriate, targeted recommendations for diverse communities. Critically, the use of aggregate data as a proxy for subsets of Asian American patients obscures differences in patient- and community-level experiences or needs and conflates the experiences of minority communities within that population. This problematic practice has been well documented, and efforts are in place to address it in research and clinical practice [27,55-57]. Health informaticists and technology researchers can be change leaders in this area by applying well-established design practices such as user stories, personas, and customer segmentation to clearly identify the needs of patient users, including those that are defined by a specific cultural identity or intersections of identities [58,59].

There are several limitations to this study. We included only major databases (PubMed and Embase) and did not include unpublished or gray literature. We also limited our inclusion criteria to articles published only in English, excluding Chinese language biomedical databases such as the China National Knowledge Infrastructure. We further included only those articles focusing on populations in the United States. These criteria were established to ensure a focused review of our target community of interest, namely Chinese Americans, engaging with relatively similar health care delivery models and HIT technology. However, this may have resulted in the exclusion of relevant articles, particularly those published in Chinese language journals. Additionally, although the term “patient portal” included in our search string is broadly used, our search may have missed studies that incorporated portals, portal-like systems (eg, PHRs), or portal features without explicitly identifying them. We attempted to address this by performing a series of web-based searches (Google) and manual searches to identify articles using variable terms that could meet our inclusion criteria. Finally, our study did not systematically evaluate the quality of the data presented in the included studies beyond an assessment of the study design and the level of racial or ethnic granularity among Asian Americans; moreover, we did not evaluate the bias in these studies. Future areas of research may include expanded language contexts and further quality and bias evaluations.

Conclusions
There is limited research dedicated to understanding the use patterns, experiences, and needs of Chinese Americans who access and use patient portal systems for their health care. Most of the research in this area focuses on disparities in use and access across the aggregated racial and ethnic demographic of Asian Americans, potentially obscuring important differences among and between the diverse and heterogeneous populations that comprise this designation. Studies are also overwhelmingly quantitative, focused on surveys and administrative data from portal systems, and they lack longitudinal data. Future research should focus specifically on Chinese Americans and prioritize performing more detailed longitudinal and qualitative evaluations to understand why specific communities of patients access and use portals in the ways that they do. A broader understanding of the diversity of health technology users in general can help ensure that these tools are applicable and acceptable to all patients, including the most vulnerable, and do not contribute to disparities in health access, equity, or outcomes.

Acknowledgments
This publication is supported by the National Institute on Minority Health and Health Disparities-National Institutes of Health (grant U54MD000538). The views expressed are those of the authors and do not necessarily represent the official position of the funding organization.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Ovid MEDLINE search strategy.[PNG File , 282 KB - humanfactors_v9i2e27924_app1.png ]

References


32. Better systematic review management. Covidence. URL: https://www.covidence.org/ [accessed 2021-04-10]


44. Miles RC, Hippe DS, Elmore JG, Wang CL, Payne TH, Lee CI. Patient access to online radiology reports: frequency and sociodemographic characteristics associated with use. Acad Radiol 2016 Sep;23(9):1162-1169. [doi: 10.1016/j.acra.2016.05.005] [Medline: 27287715]


Abbreviations

- AAPI: Asian Americans and Pacific Islanders
- DISTANCE: Diabetes Study of Northern California
- EHR: electronic health record
- HIE: health information exchange
- HIT: health information technology
- PHR: personal health record
Understanding the Potential of Mental Health Apps to Address Mental Health Needs of the Deaf and Hard of Hearing Community: Mixed Methods Study

Judith Borghouts¹, PhD; Martha Neary², MSc; Kristina Palomares², BA; Cinthia De Leon¹, MPH; Stephen M Schueller²,³, PhD; Margaret Schneider¹, PhD; Nicole Stadnick⁵,⁶,⁷, MPH, PhD; Dana B Mukamel¹, PhD; Dara H Sorkin¹, PhD; Dakota Brown⁸; Shannon McCleerey-Hooper⁸, BA; Gloria Moriarty⁹; Elizabeth V Eikey¹⁰,¹¹, PhD

¹Department of Medicine, University of California, Irvine, Irvine, CA, United States
²Department of Psychological Science, University of California, Irvine, Irvine, CA, United States
³Department of Informatics, University of California, Irvine, Irvine, CA, United States
⁴Department of Public Health, University of California, Irvine, Irvine, CA, United States
⁵Department of Psychiatry, University of California, San Diego, La Jolla, CA, United States
⁶Dissemination and Implementation Science Center, Altman Clinical and Translational Research Institute, University of California, San Diego, La Jolla, CA, United States
⁷Child and Adolescent Services Research Center, San Diego, CA, United States
⁸Riverside University Health System-Behavioral Health, Riverside, CA, United States
⁹Center on Deafness Inland Empire, Riverside, CA, United States
¹⁰Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, La Jolla, CA, United States
¹¹The Design Lab, University of California, San Diego, La Jolla, CA, United States

Corresponding Author:
Judith Borghouts, PhD
Department of Medicine
University of California, Irvine
100 Theory
Irvine, CA, 92617
United States
Phone: 1 9498240246
Email: jborghou@uci.edu

Abstract

Background: Mental health concerns are a significant issue among the deaf and hard of hearing (D/HH) community, but community members can face several unique challenges to accessing appropriate resources.

Objective: The aim of this study was to investigate the mental health needs of the D/HH community and how mental health apps may be able to support these needs.

Methods: A total of 10 members of the D/HH community participated in a focus group and survey to provide their perspectives and experiences. Participants were members of the Center on Deafness Inland Empire team, which comprises people with lived experience as members of and advocates for the D/HH community.

Results: Findings identified a spectrum of needs for mental health apps, including offering American Sign Language and English support, increased education of mental health to reduce stigma around mental health, direct communication with a Deaf worker, and apps that are accessible to a range of community members in terms of culture, resources required, and location.

Conclusions: These findings can inform the development of digital mental health resources and outreach strategies that are appropriate for the D/HH community.

(JMIR Hum Factors 2022;9(2):e35641) doi:10.2196/35641
mental health; deaf and hard of hearing community; mHealth; digital health; needs assessment; deaf; hard of hearing; hearing; focus group; survey; mixed methods; intervention; health app; user needs

Introduction

Accessing mental health services is a challenge in the United States, a challenge that is further magnified for persons who are deaf and hard of hearing (D/HH). D/HH is an umbrella term used to encompass a diverse community. Other terms used by members of the community may include “deaf,” “Deaf,” or “late-deafened.” Following feedback on terminology from our participants, we have chosen to use the term D/HH throughout this paper to refer to this community, and we acknowledge that participants may use different terms to self-identify. Debate exists within the community over Deafness or deafness as disability versus Deafness or deafness as linguistic minority. While this is beyond the scope of this paper, we encourage readers to see Skelton and Valentine for an overview [1]. The D/HH community, which has often been referred to as an “invisible minority” [2], is a community with its own unique culture, traditions, and challenges. Members of the D/HH community may face significant psychosocial challenges and environmental adversity as they navigate an inherently ableist, hearing world [1,3]. Several studies do show that members of the D/HH community experience higher rates of psychological distress [4-6].

When it comes to accessing care, the D/HH community faces significant health care marginalization and health care inequities [3]. D/HH individuals report a lack of availability of mental health services [7], and Critchfield [8] estimates that 80% to 90% of people who are D/HH with severe and persistent mental illness do not receive care. As summarized by Pertz et al [3], this lack of mental health care access is multifaceted and largely stems from systemic barriers facing the community, for example, insurance coverage [9], lack of interpreters for health care visits [10,11], and lack of evidence-based, culturally competent mental health treatment options [12]. Patient outcomes for D/HH persons are better when they receive care from providers who understand Deaf culture, but these are rarely available [13,14]. Pertz et al [3] found that Deaf signers at an integrated medical and behavioral health program with a telemental health (TMH) intervention reported significantly lower depression and anxiety scores from baseline and high satisfaction ratings due to accessible communication and optional ongoing care through a TMH platform. Negative experiences and challenges communicating with ineffective providers can impact treatment engagement and adherence [15], creating general distrust, reluctance, or resistance to the mental health care system [16].

Technology provides opportunities to overcome some of the barriers to accessing mental health care traditionally facing the D/HH community. Many people in the D/HH community report using technology in other aspects of their lives. Examples include text-to-speech apps or smartphone features, such as Ava or Siri; videoconferencing, which is commonly referred to as videophoning in the D/HH community; sound enhancement apps, such as Sound Amplifier; and a variety of visual alert assistive technologies [17]. A national survey by Maiorana-Basas and Pagliaro [18] suggests that technologies such as texting, emailing, and instant messaging are used at similar rates across the population, regardless of hearing status. The rapid development of technology has led to a proliferation of digital resources designed to support and help people manage their health, and TMH services are effective treatment modalities among the general population (see Langarizadeh et al [19] for a review). TMH services are especially suited in the treatment of D/HH persons because the D/HH community may already have a level of familiarity with visually oriented technologies and assistive technologies, which may help facilitate treatment delivery [20]. TMH may also help facilitate service delivery to D/HH individuals who may otherwise not have a local, culturally competent mental health provider from whom to seek treatment [16]. Furthermore, smartphone access in the United States is increasing, although it is not ubiquitous, and several socioeconomic factors influence access, particularly considering that the average price of a smartphone is now over US $500 [21]. Those with technology access and digital literacy skills are likely to be younger, highly educated, and possess adequate financial resources. Although technology is often posited as the “great equalizer,” it can also serve to further widen the gaps between privileged and underprivileged groups, who differ in their access to, knowledge of, and ability to make full use of the medium [22].

Mobile health (mHealth) apps for the D/HH community exist, though these largely serve as assistive technologies that aim to augment people’s ability to navigate and communicate in public and with family and connect with other members of the community. In Romero et al’s [23] review of existing mHealth apps for the D/HH community, only two apps from an initial search list of 217 apps were related to mental health. They note that the relatively low yield and high turnover of available apps necessitates more development of apps for the D/HH population. There are no studies, of which we are aware, that have explored mental health apps specifically. Indeed, in our own searches of available resources to inform the development of this study’s methodology, we did not identify any apps to support the mental health of the D/HH community.

In general, while previous studies have identified several challenges among the D/HH community to access mental health resources, it is less understood if and how mental health apps may overcome these challenges. The aims of this study were to explore the mental health needs of the D/HH community and explore how digital resources such as apps may be able to support these needs. To address these aims, we conducted a focus group with 10 community members to get an in-depth understanding of their experiences and perspectives.

Methods

Overview

A community-based participatory approach was used throughout our study to engage community members in multiple stages of
the study. The effort ensured that the data collection content and processes were appropriate, the study design was suitable, and the voices of community members were accurately represented in reporting our findings.

Participants
A total of 10 people participated in one focus group, and 9 of these participants also completed a follow-up survey. Participants were members of the Center on Deafness Inland Empire (CODIE) team and based in Riverside County, California. The CODIE team comprises people with lived experience as members of the D/HH community. CODIE works to advocate for the community by empowering individuals with information, offering training and opportunities, and working to resolve challenges in areas such as communication barriers, peer counseling, independent living skills, community education, and outreach. Participants were invited by email to participate in the focus group by a lead advocate on the CODIE team.

Demographic information was collected using a web-based English-written survey distributed after the focus group. All participants reported comfort with written English, and the survey was developed in partnership with the CODIE team and Riverside County. One participant did not complete the survey, so demographic details describe 9 participants. Given the small sample size, we report the general characteristics of the sample. Participants ranged in age from 30 to 60 years (mean 44.1, SD 11.3). Participants reported their gender as female and identified as White, Black or African American, Asian, American Indian or Alaska Native, or Mexican, or they identified with more than one race. Out of 9 participants, 8 (89%) most often used American Sign Language (ASL) at home, and 7 (78%) participants reported their preferred communication method as ASL.

Measures
For the focus group, the research team developed a focus group facilitator guide with discussion topics and sample questions. Topics and questions were developed and refined based on research partners’ interests and their past learnings working with the D/HH community. The research partners consisted of staff from Riverside County Behavioral Health, peer specialists, and the lead advocate on the CODIE team. The research team met with evaluation staff from Riverside County Behavioral Health prior to the focus group to review the questions included in the guide, obtain input on the topics covered, and ensure language used was appropriate and understandable. The lead advocate also provided best practices for facilitating focus group discussions with the D/HH community. First, it was important for the facilitator to have a clear video picture in a well-lit room, tie long hair back, and minimize distractions such as moving objects in the background, so that participants could focus on body language, facial expressions, and lip movements. Second, it was advised for the facilitator to look directly into the camera and speak slowly and clearly to allow for lip reading and interpretation. Third, the facilitator should pause after asking a question to allow for interpretation and look at the interpreter to ensure that interpretation had occurred. Lastly, interpreters should introduce themselves at the beginning of the focus group and provide guidance for participants to pin them on their screen.

While these practices were given specifically for a virtual focus group, many of them are applicable for in-person focus groups too, such as speaking slowly and creating pauses for interpretation (see Balch and Mertens [24] for further lessons learned from D/HH participants on conducting focus groups).

The main focus of this study was on understanding the mental health needs of the D/HH community and how mental health apps may support the D/HH community’s mental health needs. Therefore, topics included perspectives on both mental health in general and mental health technologies specifically. Topics covered in the focus group guide included the following:

- Perspectives on mental health within the D/HH community
- Mental health needs and services available for the D/HH community
- Use of and attitudes toward apps and technologies for mental health within the D/HH community
- Challenges and facilitators to using mental health apps and technologies by the D/HH community.

The follow-up written survey was sent 5 days after the focus group. The survey asked additional questions around digital mental health and was intended to supplement findings from the focus group as well as allow participants to express thoughts outside of the focus group setting. The survey questions were developed before the focus group but were refined based on information obtained in the focus group. For example, one survey question asked what aspects of mental health apps were important to participants; the answer options of this question were updated to include certain aspects mentioned during the focus group. The four topics covered in the survey are discussed next.

The first topic was “barriers to mental health resources.” Participants were asked to report all barriers, if any, they faced to accessing mental health–related resources. They were instructed to “select all that apply” from a list of options, type free text, or both. The list of barrier options was adapted from the Healthy Minds Study, an annual web-based survey assessing mental health and service use among college students [25].

The second topic was “important aspects about mental health apps.” Participants were asked to rate the extent to which different aspects of mental health apps were important to them (eg, “The app is free”). They were asked to rate items on a scale from “not at all important” (1) to “extremely important” (5).

The third topic was “mental health app use.” A single question was used to identify whether participants had used mental health apps. In the survey, a mental health app was defined as “an application on your mobile phone or tablet device that helps you manage your mental, emotional, or psychological health.” Participants could select whether they had used apps in the past, were currently using apps, had never used apps or would be interested, or had never used apps and were not interested. Participants were also asked to rate three statements related to whether they had the resources required to use mental health apps (eg, “I have the resources necessary to use mental health apps”). The scale ranged from “strongly disagree” (1) to “strongly agree” (5). The items were based on
the facilitating conditions subscale of the unified theory of acceptance and use of technology questionnaire [26]; they were adapted to refer to mental health apps specifically.

The fourth topic was “current and desired resources to support mental health.” Participants were asked to select what resources they currently used and what strategies they wished to use to support their mental health, if any (eg, “informal support, such as talking with or spending time with family or friends”). They were instructed to “select all that apply” from a list of options, type free text, or both.

The complete survey instrument is included in Multimedia Appendix 1.

Procedure
The focus group took place on September 11, 2020, and survey data collection took place between September 16 and 28, 2020. The focus group was held online via Zoom (Zoom Video Communications, Inc), facilitated by hearing research staff, and supported by two interpreters to translate spoken English into ASL and vice versa. Each focus group question was also shown in written English in the chat window of the Zoom session, and participants were able to provide written responses in the chat window. The focus group discussion was audio recorded. The audio recording captured both the ASL translated to spoken English and the chat messages, which were read aloud. The duration of the focus group was 2 hours. The survey was distributed via Qualtrics and took approximately 20 minutes to complete. Participants received a US $30 gift card for their participation in the focus group and a US $10 gift card for completing the survey.

Ethical Considerations
The study was approved by the University of California, Irvine, Institutional Review Board (IRB; review number No. 20195406). Prior to the focus group, participants were emailed a study information sheet that was reviewed and approved by the CODIE lead advocate and the IRB. The sheet was then reviewed in the focus group session, with an opportunity for participants to ask questions. Participants were asked for their permission to audio record the conversation at the start of the focus group.

Analysis
The audio recording of the focus group was transcribed. The analytical framework used to analyze the transcript was the six-phased approach of thematic analysis as described in Braun and Clarke [27], which involves the following: (1) familiarizing yourself with the data, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) writing up the analysis. We adopted an interpretivist epistemological position and used an inductive analytical framework used to analyze the transcript was the non-D/HH researchers, who is a trained PhD researcher with transcript. The initial coding (phases 1-3) was done by one of Scientific Software Development GmbH) was used to code the transcript. The initial coding (phases 1-3) was done by one of the non-D/HH researchers, who is a trained PhD researcher with expertise in user experience and thematic analysis. For phase 4, a preliminary summary of findings was shared with study participants and other members of the D/HH community and research team who attended the focus group in order to check that this was what was said, to corroborate, to correct or extend interpretations of findings, and to further refine themes. These findings were discussed over email and during a video call meeting, where an interpreter was present to support the discussion. For phase 5, themes were defined and named by the non-D/HH research team. For phase 6, a draft of the write-up was shared with members of the D/HH community and research team who attended the focus group to provide feedback, craft the language, and add details.

We analyzed the survey data using descriptive statistics in the form of the number of people who selected certain answers. The statistical software SAS (version 9.4; SAS Institute Inc) [28] was used for analysis of the survey data. The main purpose of the survey was to supplement findings from the focus group and describe the study sample (eg, demographic information, the number of participants who had used mental health apps before, and the number of participants who wanted to use certain mental health app features that were discussed during the focus group).

Results
The following section presents an overview of study results. Unless otherwise specified, results are based on the focus group. Illustrative quotes are provided for each theme.

Current and Desired Strategies to Support Mental Health
Participants had not used any mental health apps before. A total of 6 out of 9 (67%) participants indicated on the survey that they were interested in using one, and 3 (33%) participants indicated they were not interested. Though mental health apps were not commonly used, participants shared that they used other online resources to support mental health, such as spiritual classes, meditation, and ASL yoga:

I use deaf spirituality. That kind of covers a lot of different things. Healing, holistic healing. There’s meditation. I use a website as well as a good resource, oh, to find practitioners who use sign language for all of those types of things.

Other apps that were reported to be used were communication apps to connect with others, such as WhatsApp, Zoom, and Skype.

Participants were asked on the survey what their current and desired strategies were to support their mental health. The most common strategies currently used involved informal support connecting with friends or family (n=6, 67%), peer support (n=6, 67%), and use of social media (n=5, 56%). The most common desired resource was professional mental health services (n=6, 67%), followed by activities like writing, painting, and playing or making music (n=5, 56%); online forums or communities (n=4, 44%); and websites (n=4, 44%). A total of 3 (33%) participants reported wanting to use online chat and peer support, as well as exercise programs or activities to manage their mental health.
Participants were also asked on the survey about what the broad D/HH community would like to be able to do with mental health resources. They felt the D/HH community would be most interested in resources that allow them to talk with other people to give and get support (n=6, 67%) as well as those that allow them to express themselves or have an outlet through art, photos, or writing (n=6, 67%). A total of 5 (56%) participants also reported several other possible interests, such as identifying or recognizing symptoms, working through negative emotions and thoughts, connecting with a professional, and getting information about how to cope with stress, grief and loss, trauma, and relationship issues.

**Challenges to Using Mental Health Apps**

**Support for a Spectrum of Language and Linguistic Needs Within the Community**

The most reported barrier to accessing mental health services on the survey and during the focus group was the difficulty in finding mental health care providers that knew ASL. Similarly, the main barrier to using online mental health tools specifically, as reported on the survey, was difficulty finding a tool that supported ASL.

Beyond a lack of ASL support, participants reported issues to accessing mental health services with respect to communication, access, and feeling welcome. Participants shared that there are a range of language and linguistic needs within the community, with some people feeling more comfortable with English, whereas others are more comfortable with ASL. Furthermore, there are different literacy levels within the community in terms of understanding English. Participants recommended providing different options to present content through a digital mental health intervention, such as text, videos, and icons, and providing ASL video where possible. One participant noted the following:

_Talking about English and ASL, there’s neither one that is better than the other. It’s a matter of what the person feels most comfortable with...You also don’t want it to be just ASL only, it might force somebody out of their comfort zone. So we need to consider that spectrum of language and linguistic needs and comfort levels, which is really wide._

In the context of differing linguistic needs, participants also discussed knowledge gaps in relation to mental health concerns. The D/HH community misses incidental learning opportunities around mental health, which happen when people gain knowledge from informal interactions and overhearing conversations that can be related to societal changes in attitude toward mental health. These learning opportunities typically rely on spoken language.

**Lack of Accessible Services**

Even if an interpreter could be provided to aid communication with a mental health provider, participants reported that many community members would not feel comfortable with having an interpreter present and would feel safer if they could speak directly to a mental health provider with the same language. One participant explained the following:

_Sometimes interpreters will use different word choices and it’s not what I mean. Or you know, confidentiality, because people may want to keep that privacy._

One participant mentioned that to build a connection with a health provider, it helps to talk to someone who looks and signs like them. An additional barrier to accessing mental health services was that providers were not sensitive enough to cultural differences. For example, a participant explained that hearing providers do not have a “deaf heart” and the sensitivity or the same experiences as them.

Participants said there was a lack of Deaf workers in the mental health profession and that it was challenging to find mental health services for the D/HH community. Participants expressed concerns that there was a lack of accessible resources overall and for specific services, such as marriage counseling, anger management, substance abuse treatment, and support for domestic violence. For services that were available, participants said that community members were sometimes limited in terms of their insurance and what services they could access. For example, one participant commented that services may only be available out of their state and, thus, not covered by insurance, or that the only ASL services available are very basic.

**Stigma Around Mental Health**

There was a consensus during the focus group that stigma around mental health was a considerable challenge in the community, and that community members did not want other people to know they were accessing mental health services. Participants had concerns over the use of the term “mental health” and said that positive and uplifting terms centered around spirituality and healing would be more appropriate, stating that these would resonate more with the community and signal that a positive experience is forthcoming. For example, one participant suggested the name “Healing Hands” for a mental health app.

Participants expressed concerns that for many community members, miscommunication has had negative repercussions in the past, which can increase stigma, and there are fears of experiencing negative side effects of getting treatment. For example, members with children may have fears that if they talk about their mental health challenges, their family, such as their children, may be taken away or there may be financial consequences. One participant stated that mental health can be perceived as “just another thing wrong with my head.”

Participants noted that community members may have privacy concerns around the use of mental health services, such as concerns around whether their information was going to be safe. On the survey, 3 (33%) participants indicated that they had privacy concerns on their personal information being visible by using mental health apps.

Participants expressed a need for increased education and awareness around mental health, and to promote a message that mental health services are helpful in a good way, that it was okay to seek help, and that mental health is for everyone:

_Nowadays what I’m seeing is Deaf and Hard of Hearing people putting vlogs emphasizing it’s okay_
to feel whatever you’re feeling, and it’s okay to look for help. And I think that’s key, if developing an app...to emphasize that. That’s what most of society is doing at this moment.

Facilitators to Engagement With Mental Health Apps

Overview

Participants gave several recommendations on marketing a mental health app to the D/HH community. Examples included using posters and signs, scrolling and video advertisements for the app at medical offices and social service offices, contacting nonprofit organizations that service the D/HH community, and word of mouth. It was important that the marketing materials supported a feeling of being welcome, for example, through visual advertisements that showed the step-by-step process of using the app. Participants preferred the app to be advertised with ASL people signing, using more visuals than words. Participants pointed out that instead of an “interpreter” sign, a better sign would be the two-handed sign for “peer,” “advocate,” or “support,” ideally with hands of different colors and genders. They also approved of the “same same” sign.

With respect to speaking with a health provider, participants expressed a need to be able to choose a specific person with whom they felt comfortable talking. Some people may have experienced trauma with past providers and wanted to talk to someone who would be a good fit for them regarding language and other characteristics, like gender. It was important to have diversity within the community and presented on an app, in order to make the app accessible to everyone.

Lastly, if a mental health app were to be developed that was inclusive of the D/HH community, participants expressed a preference for an app that would be useful to everyone, not just members of the D/HH community. Clicking on an app that would be specifically labeled for the D/HH community could give a feeling of being singled out. As one participant explained:

We want to try and keep that general to have access to things instead of feeling like, ‘oh okay, I have to click on this because it says deaf,’ that singles me out.

Immediate and Continuous Access to Resources

Participants placed importance on the fact that a mental health app should be accessible to a range of people in terms of language, culture, resources required to use the app, and location. They explained that community members may have limited data or memory on their phone, no access to high-speed internet, or no access to a computer. A mobile app was the preferred platform to enable people to access resources on the go. On the survey, 5 (56%) participants indicated that they had concerns about their mobile data plan when using their mobile device, 4 (44%) participants did not have the necessary resources to use mental health apps, and 3 (33%) participants were concerned about having enough space to download apps on their smartphone.

Participants also expressed value for an app to provide immediate access to resources and services. At the time of the focus group, people had to go through a long intake process before they could connect to mental health services:

Having that immediate assistance, to somebody live or whatever it is, right there is important rather than having to go through all of these different things...you have to go through all of that demographic information and you have to basically tell your life story before you can get to somebody.

Given the range of literacy levels within the D/HH community, participants worried that people may not understand all intake questions, which can slow down the process further and reduce interest in engagement. A participant mentioned that having the intake available in both English and ASL would likely make the intake process go smoother.

Participants also said that it would be ideal to have unlimited access to resources, as opposed to there being a limit to the number of times they could access them. Especially during the global COVID-19 pandemic, services were sometimes used by people just to connect and talk to someone, and some consumers accessed mental health services multiple times a day.

Though participants were part of a local community-based agency, they named several benefits of making an app globally accessible to anyone, rather than tying it to a specific location. First, people may be located elsewhere but prefer coming to a specific organization, such as CODIE, for services. Second, there were benefits to working together with other organizations. If someone is in need outside of standard business hours, there may not be anyone near them to help, but there may be someone awake in another time zone and location who can provide support. Third, these collaborations could facilitate linking to resources from other organizations, in order to spread awareness and help the broader D/HH community. Participants suggested making apps available to people from Gallaudet University, the only university in the world where students learn in ASL and English.

Important Features for Mental Health Apps

Participants were asked on the survey what the most important aspects were that mental health apps should offer. The most important aspects were suicide prevention support, emergency support, peer support and chat, and telehealth (ie, referring to a direct connection to clinical mental health services within the app). A chatbot was rated as the least important.

During the focus group, participants reported that it was important for people to interact with a human and not an avatar. An avatar is a computerized figure, such as an icon, that can represent a simplified figure of a person. The term avatar can refer to different things and can also be used for digitally created characters that turn speech into sign language. Participants’ reservations about an avatar were that it lacked facial expressions and body language, which are important to have in addition to signs in communication. One participant explained as follows:

We want an actual person, not an avatar because an avatar lacks that body language, those expressions that, a lot of people that maybe are at the gestural language level would understand the body language more than the actual signs.
Furthermore, they reported that it was valuable to build a real relationship with a human being. It was not discussed during the focus group whether participants did not want avatars at all or whether avatars could be used to supplement interaction in an app, if it also allowed connection with a live person.

Participants mentioned past use of Deaf clubs, which are places where Deaf people can meet face-to-face and socialize [29]. Participants shared that they used Deaf clubs for socializing, fun, and games, and expressed the desire to see these Deaf clubs being used for mental health support. It was discussed whether an app could have something similar to a Deaf club.

Inclusion of Diverse Community Members in Technology Development

To reduce stigma around seeking help for mental health, participants suggested having members of the community contribute to an app, for example, through blogs or short videos to share their experiences and knowledge. Participants liked the aspect of inclusivity where visitors can be “part of the news”:

Like a blog that people could add—share their experiences and their knowledge and their education...They could be a part of the news basically. Instead of watching the news with captions, they could watch this person signing that.

In addition to providing content, participants recommended having community members involved in the design process to give feedback on features and on what can be improved. Some participants reported that it would also be valuable to have community members provide guidance on how to access and use the app, for example, through instruction videos and visual posters with step-by-step instructions.

Discussion

Principal Findings

The aim of this paper was to understand the mental health needs of the D/HH community and how mental health apps may be able to support these needs. In line with previous work [7,9], participants indicated that community members had limited access to mental health resources. Digital solutions, such as mental health apps, may increase access to resources, but our study highlighted that it is important to take certain factors into consideration to facilitate engagement with such apps.

Some of the themes we found are common barriers among other mental health–seeking populations as well, such as stigma [30]. This barrier may be exacerbated in the D/HH community due to missed incidental learning opportunities about mental health. Furthermore, similar to previous work with hearing populations [31], participants valued immediate access to resources. Participants rated access to suicide prevention support and peer chat as one of the most important features to include in a mental health app, which resonates with work with other communities: a recent study with essential workers found that one of the most desired features for a mental health technology was the ability to chat with a mental health professional or peer, and a link to mental health resources and crisis support [31].

Factors that may be more unique to the D/HH community are the need for both ASL and English support, and the finding that participants wanted a general app that is inclusive of the D/HH community, rather than an app exclusively made for them. For example, participants emphasized the importance of including members of the D/HH community on the app, but to market an app as usable for everyone to avoid singling out the D/HH community. This finding further supports the need for customization and personalization of mental health apps [32,33] and the importance of inclusive design and designing for a wider population [34]. The ability to customize an app to a user’s personal needs can facilitate feelings of perceived fit to a user’s culture and values [35], without singling out a particular community.

While further follow-up studies are recommended to corroborate themes with a larger population, initial takeaways can be extracted from our findings to inform creation and development of digital mental health resources, such as apps. Below we outline several learnings that may be important to consider when developing digital mental health resources for the D/HH population.

Support for a Spectrum of Language and Linguistic Needs Within the Community

Similar to previous work [10,36], the greatest barrier to accessing mental health services identified by participants pertained to communication issues. Participants reported a lack of Deaf workers and mental health care providers that knew ASL. It is, thus, important to support a spectrum of linguistic needs within the community.

Participants in our focus group primarily highlighted the limitations of English-language mental health services, rather than positive experiences. A previous focus group with D/HH community members indicated that there may be social pressure during ASL focus groups that limits participants from sharing any positive experiences with English health care communication [37]. While our study participants acknowledged that some community members may prefer English, the main issue was that there needs to be support for a range of linguistic needs, rather than English or ASL support alone. Participants shared that there are various language and linguistic needs within the community, with some people feeling more comfortable with English and ASL, and there are different literacy levels in terms of understanding English. Our survey further found that a lack of Deaf workers in the mental health profession was not the only barrier to accessing resources, but it was the most common.

While previous studies found difficulties in accessing mental health resources due to a lack of interpreters [37,38], participants in our study indicated that the availability of an interpreter alone may not be sufficient. Even with an interpreter present, participants explained that community members may not feel comfortable talking via an interpreter and prefer to communicate with a Deaf worker directly. This sentiment is consistent with findings from Steinberg et al [39] who found a preference for health care practitioners who are fluent in ASL and support from other Deaf individuals.
To support a range of linguistic needs, participants recommended providing different options for presenting content through an app, such as text, videos, and icons, and providing ASL video where possible. Participants also provided marketing suggestions to support a feeling of being welcome, for example, through visual advertisements that show the step-by-step process of using the app.

**Stigma and Appropriate Use of Terminology**

Similar to prior studies with D/HH community members [39], study participants expressed that there was stigma within the community around mental health issues and seeking help for these issues. Though stigma can be a common barrier among the population in general [30], it may especially be an issue for the D/HH population, as they are not exposed to mental health issues and information in the same way as the general public [38]. Even though study participants expressed a need for more Deaf workers in the mental health profession, previous work found that because of this lack of exposure, Deaf workers may be less knowledgeable about mental health issues [38].

Prior work on mental health apps has suggested that delivering support through technology can overcome stigma barriers, as people do not need to know one is seeking help [40]. However, participants in our study still had concerns about their information not being private through an app. To help mitigate privacy concerns, it is important to be transparent on how app data are collected and stored and how they will be used. Furthermore, participants recommended that instead of using the term mental health, positive and uplifting terms around healing are preferred in order to facilitate adoption of a mental health app.

**Education Around Mental Health**

Communication issues can complicate accurate reporting of mental health prevalence in the D/HH community [41,42], and our findings further showed that D/HH members can often miss out on informal conversations and may not be as knowledgeable about mental health as the hearing population. This finding highlights that increased education around mental health may be especially important for this community. Participants expressed a need for increased education and awareness around mental health, for example, through short videos and by having members of the community share their experiences. Participants stated that there was a need for people to understand that mental health services can be helpful, and that strong mental health is a goal for everyone.

**Include Members of the D/HH Community and Market for Broader Community**

Ideally, participants preferred to have direct communication with a Deaf worker that had the sensitivity and experience to communicate with members of the community. Participants also recommended involving community members in providing content and sharing feedback about improving app features. It was important to have an app that is inclusive of, but not exclusively for, the D/HH community. Participants preferred an app that would be useful for anyone and that would not just be focused on their community, which may exacerbate feelings of being singled out.

**Limitations and Future Work**

The study has several limitations. First, care should be taken to generalize its findings to the broader community. An advantage of a focus group setting is that it has been shown to be a suitable methodology for Deaf culture to gather and share information in a safe setting [37,38], but sample size is limited. In addition, participants in the focus group were engaged with an advocacy group and involved in the community, so their experiences may be different than those of the general community. Participants had experienced hearing loss since birth or early in life, and their experiences may differ from those who experience hearing loss because of old age or those who experience hearing loss later in life. Our study offers insights into how mental health needs of the D/HH community may be supported through digital therapeutics, which would be worthwhile to explore further in a larger-scale study. Second, an English written survey was used to collect participants’ demographic information. Though all participants were able to read and write in English, the majority of participants’ preferred language was ASL, and one participant did not complete the survey. Our study results have since been used to inform a collaborative effort to create an ASL survey for broader needs assessment with the D/HH community. Third, results were collected during the COVID-19 pandemic, which may have increased mental health concerns and interest in mental health resources. Lastly, we used the English translation of the focus group discussion for data analysis. There may be limitations in using an English translation, as information may be filtered and expressions can differ from ASL. For example, personal pronouns in ASL are not gender specific [37]. To ensure that our analysis and findings accurately represented participants’ views, we refined our findings through member checking with the focus group participants who are members of the D/HH community.

**Conclusions**

This study looked at the mental health needs of the D/HH community and how mental health apps may be able to support these needs. There was a need for more Deaf workers and ASL support to support a spectrum of linguistic needs; a need for increased education to reduce stigma around mental health; a need for an app that is accessible to a range of people in terms of culture, resources required, and location; and a need for immediate and unlimited access to resources. These findings are important to consider for the development and dissemination of mental health apps to meet the needs of the D/HH community.

**Acknowledgments**

We deeply appreciate the members of CODIE who shared their experiences and feedback with us and were patient as we learned more about the D/HH community. Without their expertise, these insights would not be possible. We would also like to thank the
interpreters as well as Suzanna Juarez-Williamson, Riverside County, the Peer Support Team of the Help@Hand Project, and the Help@Hand Evaluation Team.

This work was funded by Help@Hand (contract No. 417-ITSUCI-2019), a project overseen by the California Mental Health Service Authority (CalMHSA). CalMHSA reviewed the manuscript for confidentiality. The information or content and conclusions presented here are those of the authors and should not be construed as official position or policy of, nor should any endorsements be inferred by, the participating Help@Hand counties or CalMHSA.

**Conflicts of Interest**

SMS has received consulting payments from Otsuka Pharmaceuticals and Trusst (K Health) and is a member of the Headspace Scientific Advisory Board, for which he receives compensation.

Multimedia Appendix 1

Survey items.

[DOCX File .41 KB - humanfactors_v9i2e35641_app1.docx ]

**References**


Abbreviations

ASL: American Sign Language  
CalMHSA: California Mental Health Service Authority  
CODIE: Center on Deafness Inland Empire  
D/HH: deaf and hard of hearing  
IRB: Institutional Review Board  
mHealth: mobile health  
TMH: telemental health
Process and Information Needs When Searching for and Selecting Apps for Smoking Cessation: Qualitative Study Using Contextual Inquiry

Ylva Hendriks¹, MSc, MPIM; Sebastiaan Peek¹, PhD; Maurits Kaptein², PDEng, PhD; Inge Bongers¹,³, PhD

¹Tranzo, Tilburg School of Social and Behavioral Sciences, Tilburg University, Tilburg, Netherlands
²Jheronimus Academy of Data Science, ’s-Hertogenbosch, Netherlands
³Research Unit Evidence Based Management of Innovation, Mental Health Care Institute Eindhoven, Eindhoven, Netherlands

Corresponding Author:
Ylva Hendriks, MSc, MPIM
Tranzo, Tilburg School of Social and Behavioral Sciences
Tilburg University
PO Box 90153
Tilburg, 5000 LE
Netherlands
Phone: 31 13 466 4892
Email: ylva.hendriks@tilburguniversity.edu

Abstract

Background: Hundreds of apps are available to support people in their quest to quit smoking. It has been hypothesized that selecting an app from a sizable volume without any aid can be overwhelming and difficult. However, little is known about how people choose apps for smoking cessation and what exactly people want to know about an app before choosing to install it. Understanding the decision-making process may ultimately be helpful in creating tools to help people meaningfully select apps.

Objective: The aim of this study is to obtain insights into the process of searching and selecting mobile apps for smoking cessation and map the range of actions and the accompanying reasons during the search, focusing on the information needs and experiences of those who aim to find an app.

Methods: Contextual inquiries were conducted with 10 Dutch adults wanting to quit smoking by using an app. During the inquiries, we observed people as they chose an app. In addition, 2 weeks later, there was a short semistructured follow-up interview over the phone. Through convenience and purposive sampling, we included participants differing in gender, age, and educational level. We used thematic analysis to analyze the transcribed interviews and leveraged a combination of video and audio recordings to understand what is involved in searching and selecting apps for smoking cessation.

Results: The process of finding smoking cessation apps is comprehensive: participants explored, evaluated, and searched for information; imagined using functions; compared apps; assessed the trustworthiness of apps and information; and made several decisions while navigating the internet and app stores. During the search, the participants gained knowledge of apps and developed clearer ideas about their wishes and requirements. Confidence and trust in these apps to help quitting remained quite low or even decreased. Although the process was predominantly a positive experience, the whole process took time and energy and caused negative emotions such as frustration and disappointment for some participants. In addition, without the participants realizing it, errors in information processing occurred, which affected the choices they made. All participants chose an app with the explicit intention of using it. After 2 weeks, of the 10 participants, 6 had used the app, of whom only 1 extensively.

Conclusions: Finding an app in the current app stores that contains functions and features expected to help in quitting smoking takes considerable time and energy, can be a negative experience, and is prone to errors in information processing that affect decision-making. Therefore, we advise the further development of decision aids, such as advanced filters, recommender systems and curated health app portals, and make a number of concrete recommendations for the design of such systems.

(JMIR Hum Factors 2022;9(2):e32628) doi:10.2196/32628
Introduction

Background

It is well-established that the toxins in tobacco cause a range of diseases and disorders, often leading to death [1]. The World Health Organization estimates that tobacco kills up to half of its users, which adds up to >8 million people each year [2]. In addition to its major impact on mortality worldwide, tobacco use also results in a great number of morbidities [1]. Smokers die younger, age more quickly, and develop diseases of nonsmokers at a much younger age [3], decreasing the quality of life earlier in life.

Smoking cessation yields specific benefits of reducing fatal and nonfatal vascular, respiratory, and neoplastic (cancer) diseases [4]. Quitting cuts the risk of developing smoking-related diseases, such as lung cancer, by half [5] and increases life expectancy. Regardless of age, quitting smoking is always advantageous to one’s health. Smokers who successfully quit smoking before the age of 40 years avoid nearly all the increased mortality risks of continued smoking [4]. After the age of approximately 40 years, every year of smoking prevention saves an average of 3 months of healthy life [6]. Even stopping at the age of 60 years will gain a person 3 years of life expectancy [7].

Mobile apps, which are small software applications that run on mobile appliances, such as smartphones and tablets, are generally regarded as useful tools that aid people in their attempts to quit smoking for several reasons. For example, apps can provide highly individualized and intensive interventions [1,8-11]. Furthermore, apps have the ability to reach large audiences, which makes them cost-effective for both users and suppliers [1,8-11]. Moreover, apps can allow users to tailor interventions according to their personal needs [8]. Finally, apps can reach audiences who might not otherwise seek support [11], in part as apps allow for anonymity [12].

In addition, the persuasive technology literature shows that apps have certain characteristics that make them potentially suitable for supporting behavior change [12,13]. For instance, they can tirelessly continue to try to persuade users without getting annoyed or impatient. They are accessible at any time from any place and consequently able to support people in their behavior change even at night or in the privacy of their homes [1,8-10,12]. Furthermore, people sometimes view their smartphones as an important information source (ie, logo, title, and screenshots). However, the ratings are a reflection of app success, which is, in turn, determined by factors such as the number of languages supported, package size, app release date [25], free app offers, high volume, high user review scores, and continuous quality updates [26]. Although the provided information cues may be informative, tools to guide users through the massive number of results seem to be lacking [27]. At this moment, the visitor cannot use advanced search, filtering, or sorting options in either store. The immense supply of health apps, combined with the lack of tools for refined searching, creates a situation where choosing an app based on anything other than popularity could be considered a challenge.

Related Research

Quantitative studies on uptake, which is the act of downloading and installing smartphone apps in general, have shown that apps with a low price, high ranking, many reviews, and high ratings have the most installations [28] and that high ratings associate more strongly with downloads if customers show a degree of unanimity in their ratings [29]. This implies that these are important information cues for people when choosing apps in general. Diverse qualitative studies have confirmed these
In these studies, participants indicated that they relied on the heaviness in ratings, reviews, screenshots, and ranking when choosing various kinds of apps, including apps for smoking cessation [30-32]. Low price or the ability to try an app free of charge are important [32,33], as are the recommendations of others [33,34], preferably given by trusted sources [32].

Specifically, for smoking cessation apps, a few studies have shed light on what people consider important, desirable, or attractive features of smoking cessation apps and which functions people believe to increase engagement. Examples include ease of use, receiving feedback, goal setting, social sharing, competition, and reminders [31,33].

Owing to a recent think-aloud study [35], we now know more about potential users’ views on factors such as capability, opportunity, and motivation influencing the uptake of health apps. In this study, Szinay et al [35] found that participants considered searches for health and well-being apps to be difficult, with some calling it a minefield. Furthermore, it was shown that during the search, people pay attention to the look and design, costs, and perceived utility of apps, among others, but primarily to the opinions of others.

These studies provide clear insights into what people generally consider important about apps and which information cues people use before downloading and installing an app. Nevertheless, to the best of our knowledge, it is still unknown how all these insights come together in the process of searching for and selecting health apps in general and apps for smoking cessation in particular. As we know little about the process, we can presently only make assumptions about what the combination of the large supply and lack of tools means for people who want to choose health apps.

**Objective**

The current gap in the body of knowledge on what people do, experience, and need during the search for mobile apps for smoking cessation creates a need to better understand the process of selecting apps. Understanding the diverse information needs and decision-making processes may ultimately be helpful in creating tools to help people meaningfully select apps. What do people do and experience when searching for an app for smoking cessation? Which information is important to people when choosing an app? How do people use the available information cues in app stores (such as the Google Play Store and the Apple App Store) to obtain the desired information? This study addresses these questions by means of contextual interviews during which people choose an app for smoking cessation. This qualitative approach gives us the opportunity to elicit in situ detailed information to create a rich image based on actual behavior and people’s spoken thoughts while in action.

**Methods**

**Study Design**

Contextual inquiry is a technique for gathering field data by conducting field interviews with users and studying a task while it is performed in the everyday context. Directly observing the performance of the task enables the revelation of habitual and unconscious practices and is easier for participants as they do not have to articulate their practices [36,37]. A typical contextual interview, similar to a regular interview, begins with an introduction and some general questions about the participant’s situation and then moves on to observation of, and discussion about, the task under study. The researcher not only observes the participant’s actions but also pays attention to verbal clues and body language [37]. The distinctive characteristics of a contextual inquiry are the principles of apprenticeship and partnership. In a contextual inquiry, the researcher explicitly assumes the role of apprentice and recognizes the respondent as an expert in her or his task. Taking on this role creates a mindset that is focused on curiosity, inquiry, and learning [36]. This mindset is related to working in partnership, which facilitates true collaboration between the interviewer and the respondent to understand the task and motivation of the respondent [36]. This means that the researcher shares thoughts and confusion with the participant on the spot, thus inviting the participant to work together to understand what is happening and why.

Although contextual inquiry originates from and is typically used in contextual design projects [38-40], the method can also be applied to eHealth research [41] on, for instance, mental health [42], healthy eating [43], and persuasive technologies that facilitate healthy lifestyles [44].

**Sampling of Participants**

We recruited people who wanted to quit smoking, were interested in using an app to do so, and did not currently have or use such an app. Having used a smoking cessation app in the past was not a reason for exclusion. Additional inclusion criteria were (1) owning a mobile device, (2) knowing how to download apps, and (3) being fluent in Dutch.

We recruited participants through posters and social media and by approaching people (who were smoking cigarettes) on the streets in diverse locations in the Netherlands. In addition, we recruited participants through email within our own network. Finally, we used the snowball sampling technique by asking participants at the end of the interview whether they knew someone who might also be interested in participating. To reach our goal of understanding the diverse ways in which people search for smoking cessation apps, we purposively aimed to create variations in age, educational level, and gender.

We created a simple webpage (Qualtrics) in which those interested could leave an email address. Every channel of recruitment contained a link to this webpage. We acquired 20 leads for potential participants whom we sent an information letter. We contacted every lead after a few days to check for interest in participating in the study. Of these 20 individuals, 5 (25%) no longer reacted to our messages, and 5 (25%) had decided not to participate. The reasons stated were not wanting to quit smoking or not wanting to use an app to quit after all, no interest in participating, or practical reasons. Of the 20 individuals, 10 (50%) participated in the study. During analyses of the data, we found that we had reached saturation and, therefore, decided not to recruit additional participants (see the Strengths and Limitations section).
Procedure and Data Collection

Some weeks before each interview, we sent participants an information letter, informed them about the use of audio and video recordings, and scheduled an appointment for the interview.

**Contextual Inquiry (Interview)**

Interviews were conducted face to face (one on one) at a location chosen by the participant. Of the 10 participants, we interviewed 5 (50%) in their homes, 3 (30%) at the university, and 2 (20%) at their workplaces. No one else was present besides the participant and researcher, except for in 1 interview. Researcher SP conducted the interview, and researcher YH, who conducted the other 9 interviews, was present as an observer. Interviews were recorded using a digital voice recorder. During the search for an app, the screens of the participants’ appliances were shared with the researcher’s laptop (using Mobizen). The footage was recorded using the Microsoft PowerPoint function Insert Screen Recording. This captured both footage and sound. The researcher also took notes during the interviews to mainly facilitate revisiting certain remarks and provide a recap at the end of the interview. An interview guide was used to maintain consistency between and direction during the interviews.

Every session started with an introduction explaining the purpose of the study, talking about expectations, asking permission for recording, and answering participants’ questions. Participants subsequently provided informed consent on paper.

The introduction was followed by a semistructured interview in which we collected data on age, educational level, and smoking behavior of the participants by asking them. We also talked about prior experiences with eHealth apps, especially for smoking cessation, and about prior experiences with quitting attempts. To get a feel for the motivation of each participant to quit smoking, we used motivation rulers for smoking cessation [45]. On a scale of 0 to 10, we asked participants to indicate the extent to which they considered quitting important, how ready they felt to quit, and how confident they were about quitting. Importance, readiness, and confidence have been associated with smoking behavior change and higher scores, especially on confidence, indicating a greater likelihood of attempting to quit [45].

Subsequently, in the contextual interview, we collected data on the process of searching and selecting apps for smoking cessation. We instructed people to search for an app in the way they normally would if we were not present and gave no further instructions on where to start or how to go about the task. We told participants that the task would be completed as soon as one found an app that they considered good, adding that deciding there were no good apps and downloading nothing was also a valid option. We asked the participants to tell us aloud what they were doing, thinking, and feeling. We also asked questions about the task during the search, such as “what is your feeling, when you look at this app?” or “why did you go back to the search results?” (Multimedia Appendix 1 [36,45,46]).

After the participants made their final choice for an app, we jointly created a summary of the entire search process. Doing this together with the participant served as a means of checking our interpretations. By sharing our interpretations and being honest about interpersonal cues, we aimed to create a valid understanding [47]. In addition, questions we did not ask during the search to not interrupt the participant could be asked here. Before closing off, we informed the participants about the follow-up procedure and planned the date for a follow-up phone interview.

The length of the full sessions (from introduction to completion) varied from 50 minutes to 2 hours and 40 minutes, with an average of 1.5 hours (SD 34 minutes). The duration of the actual searches ranged from 17 minutes to 1 hour and 40 minutes (average 46, SD 26 minutes). It is important to note that this does not necessarily reflect pure search time, as, during the search, participants frequently explained their choices and voiced their ideas and thoughts. Therefore, search time is more related to the verbosity of the participant rather than to, for example, the number of apps that were reviewed.

**Follow-up Phone Interviews**

After 2 weeks from the contextual inquiry, we called the participants over the phone for a final, short semistructured interview. The researcher called the participant at the agreed-upon time. Before starting the interview, we once again asked permission to record the conversation. To do this, we used a digital voice recorder and an Olympus Telephone Pick-up Microphone. Again, we used an interview guide for the topics we wanted to address. Telephone interviews lasted between 10 and 34 (average 19, SD 9) minutes.

In the follow-up interviews, we collected data on the realization of expectations about the chosen app. Some topics we touched upon were as follows: did the participant use the app, and did the app meet the expectations of the participant, given what the participant had learned about the app during the search? In addition, we asked participants whether they had quit smoking (for topics, see Multimedia Appendix 2). Finally, we used the follow-up interview as an opportunity to come back to things participants had said or done during the contextual inquiry, which needed further clarification.

Afterward, all participants received a €15 (US $16.35) gift voucher via mail as a token of gratitude for their participation.

**Data Analysis**

Audio recordings of the interviews and the phone interviews were transcribed verbatim using the f4 transcription software. We used Microsoft PowerPoint to create the so-called process charts in which we combined corresponding screenshots and participant quotes. These visualizations enabled us to link images on the participants’ screens to what people said at that moment (Multimedia Appendix 3). A particular strength of these visualizations was the possibility of seeing that sometimes participants misread or misinterpreted information on their screens. Quantifiable information, such as the number of apps that participants looked at and the scores on the motivation rulers for smoking cessation, was transferred to Microsoft Excel sheets. From there, we translated some data into categories. For instance, we converted information about the number of cigarettes smoked per day into three categories: light smokers
(does not smoke daily), moderate smokers (<20 cigarettes/day), and heavy smokers (≥20 cigarettes/day) [48].

Textbox 1. Description of steps in data analysis.

<table>
<thead>
<tr>
<th>Stage and description</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Familiarization with the data:</strong> YH transcribed data, read and reread every transcript while listening to the recordings, and created extensive notes and memos on everything that attracted attention. We created, for instance, a memo about the observation that during the search, multiple participants wondered whether certain app features were suitable for them and whether they could see themselves using them.</td>
<td></td>
</tr>
<tr>
<td><strong>Generating initial codes:</strong> YH marked all possibly relevant text fragments to condense the data and clear out noise. In this step, YH also complemented memos and created new ones. The first transcript was coded independently by both SP and YH. The 2 versions were discussed in detail, and agreement was reached on what and how to code. A final single coded version was created. The remaining transcripts were coded by the first author (YH) while regularly conferring with the second author (SP).</td>
<td></td>
</tr>
<tr>
<td><strong>Searching for themes:</strong> YH and SP identified the initial main themes, such as starting situation of participants, navigational patterns, and use of information cues to structure the remainder of the analysis process.</td>
<td></td>
</tr>
<tr>
<td><strong>Reviewing themes:</strong> YH reviewed the initial themes by going through every transcript and process chart, one theme at a time, selecting text snippets and systematically creating headings and ordering fragments under the headings (open coding [50]).</td>
<td></td>
</tr>
<tr>
<td><strong>Defining and naming themes:</strong> In this step, to refine ideas about the themes and the narrative of the data, YH rearranged the headings, reorganized the text fragments, and reduced the number of headings (axial coding [50]).</td>
<td></td>
</tr>
<tr>
<td><strong>Producing the report:</strong> YH created the arrangement of the report using the themes on the final classification as headings. The final data analysis was interwoven with the writing process, meaning that we continuously alternated between writing, checking data, adjusting paragraphs, rearranging text, and selecting vivid and appropriate extracts to clarify the report of the results. Multiple iterations of the report were shared, discussed, and refined by all authors. For full, transparent reporting of this study, we used the Standards for Reporting Qualitative Research [51] (Multimedia Appendix 4, [51,52]).</td>
<td></td>
</tr>
</tbody>
</table>

Ethical Approval

Ethical approval for the study was obtained from the institutional review board of the YH’s university—the ethics review board of the Tilburg School of Social and Behavioral Science (reference EC-2018.92).

Results

Overview

By analyzing the data from the interviews and contextual inquiries, we identified several facets that play a role in the search for smoking cessation apps (Textbox 2). For the sake of readability and clarity, the report in the Results section is structured according to the process steps, and the themes or subthemes are addressed in the description of the process step they relate to. Furthermore, the Principal Findings section contains a descriptive overview and summary of the themes or subthemes.

The remainder of the Results section is organized as follows: we start with a description of our participants, their experience with attempts to quit smoking in the past, as well as their previous experiences with smoking cessation aids and eHealth in general. Then, we describe the identified steps of the search process and search thoroughness. Next, we describe the results per process step, focusing on, among others, participants’ information needs, actions, decisions, the reasons for those decisions, and participants’ search experience. We then describe the transformation of knowledge, wishes and requirements, and confidence in smoking cessation apps throughout the search and across the process steps.
### Major themes and subthemes

- **Search process**
  - Extensiveness and thoroughness
  - Decision moments
  - Differences and similarities between process steps

- **Information needs**
  - Information cue use
  - Functioning of apps
  - Trustworthiness and personal relevance of the information
  - Availability of information

- **Information processing and decision-making**
  - Activities, cognitive processes, and cognitive load
  - Availability of information
  - Errors in information (processing)

- **Transformations**
  - Knowledge
  - Wishes and requirements
  - Confidence in apps

### Sample Descriptive

The average age of the 10 participants was 41.2 (SD 8.7; range 26-59) years; 6 (60%) were women, and 4 (40%) were men. Of the 10 participants, 4 (40%) had higher education, 4 (40%) had middle education, and 2 (20%) had lower education. Every participant had started smoking as a teenager, at an average age of 16 (SD 1.8; range 13-18) years. This means that the participants had been smoking for 10 to 45 (mean 25, SD 9.3) years. Our sample of 10 participants comprised 4 (40%) heavy smokers, 5 (50%) moderate smokers, and 1 (10%) light smoker. Half of the participants mentioned stress relief as their main reason for smoking, and regarding *being a smoker* as something positive (self-image). Most said that they probably kept smoking as it was a habit and an addiction.

### Quitting Smoking

All 10 participants had made serious attempts to quit smoking in the past: 5 (50%) participants made one attempt, 4 (40%) participants made between 2 and 6 attempts, and 1 (10%) participant reported trying 20 times. Some memories of quitting attempts in the past were negative. For example, one of the participants recalled the freedom she felt to be independent of tobacco. Another remembered the fun, game-like aspect of no one noticing that he had quit. However, most recollections of quitting attempts in the past were negative. People remembered how hard it was to quit, how ill-tempered and irritated they felt, and the guilt and shame when the quitting attempt eventually failed. Some participants specifically mentioned losing faith in their own capability to quit and being afraid of trying again:

> *I sooooo want to quit smoking. If I had to give it a number it would be a 10, but I am terrified to fail again.* [participant 5]

Reasons for wanting to quit again were health (10/10, 100%), the sake of the children (5/10, 50%), and general negative aspects of smoking such as costs, bad smell, and social disapproval. For most participants, the *health reason* was merely a rational, calculated consideration, as most of the participants did not experience any health problems at the time of the interview:

> *Yes, you see, if I continue smoking the chance of diseases and such is big, so then...But right now I’m fit and healthy. So in the short term that is not a motivation, but in the long run it is.* [participant 7]

On the motivation rulers, the participants scored an average of 7.5 (SD 1.35; range 5-10) on the importance of quitting smoking, an average of 6.8 (SD 2.08; range 4-10) on the readiness to quit, and an average of 5.7 (SD 3.55; range 0-10) on being confident that they will quit in the next attempt.

### Experience With Smoking Cessation Aids and Apps

Almost every participant had tried some form of smoking cessation aid in the past, ranging from hypnotherapy, acupuncture, and laser therapy to medication, chewing gum, and nicotine patches. Overall, 6 participants had used a smoking cessation app on previous quitting attempts. All participants had fairly low expectations of the benefits of all these aids.
Everyone seemed to feel that quitting is something you need to do by yourself, that it is going to be hard no matter what, and that these aids can be a helping hand at most. This sentiment also applied to apps for smoking cessation. Although people found certain functions in smoking cessation apps somewhat useful or motivating, there were more comments on negative aspects, such as the inability of the app to engage them, having to pay to get access to more content, and a lack of interesting functions.

## Search and App Selection Process

### Overview

The basic steps in the search process were the same for all participants (Figure 1): every search started with entering a search query, which led to a set of results. The next step was to choose a result to obtain detailed information. Subsequently, participants decided to either return to one of the earlier steps or move on to downloading an app. Every participant opened the downloaded apps before deciding to either choose the app or continue the search. All searches ended with participants choosing at least one app they intended to use during their quit attempts.

Although every participant’s search fitted this general process, we also saw some differences. First, we could discern 2 levels of complexity in search flows. Of the 10 participants, 6 showed a simple linear flow. They went from search queries to results, inspected between 2 and 7 different detailed app information screens, and subsequently chose 1 or 2 apps to use. The remaining 4 participants showed a more complex, elaborate flow, with more loops back to the previous process steps, using more search queries, exploring more app information screens, and downloading and discarding more than one app (Table 1).

In addition to the difference in the complexity of the process flow, participants differed from each other in search thoroughness. Some participants (2/10) only scrolled a maximum of 10 apps down in the search results list, whereas other participants (3/10) examined apps in the top 20, and half (5/10) scrolled down even further, sometimes as far as 90 apps down the list. In addition, some participants (7/10) went back to an information screen they had already seen to gain new insights, whereas others (3/10) never revisited app information screens (Table 1).

![Figure 1. App selection process flow for smoking cessation apps. Thicker lines indicate more common occurrences.](https://humanfactors.jmir.org/2022/2/e32628)

<table>
<thead>
<tr>
<th>Number</th>
<th>Process flow</th>
<th>Process flow</th>
<th>App information screens (n=85), n (%)</th>
<th>Apps downloaded and opened (n=19), n (%)</th>
<th>Apps chosen for use (n=12), n (%)</th>
<th>Rank of app scrolled to</th>
<th>Revisiting information screens</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Complex</td>
<td>3</td>
<td>10 (12)</td>
<td>3 (16)</td>
<td>1 (8)</td>
<td>9</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>Linear</td>
<td>1</td>
<td>2 (2)</td>
<td>1 (5)</td>
<td>1 (8)</td>
<td>12</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>Linear</td>
<td>1</td>
<td>6 (7)</td>
<td>1 (5)</td>
<td>1 (8)</td>
<td>12</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>Linear</td>
<td>1</td>
<td>5 (6)</td>
<td>1 (5)</td>
<td>1 (8)</td>
<td>60</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>Complex</td>
<td>4</td>
<td>25 (29)</td>
<td>1 (5)</td>
<td>1 (8)</td>
<td>92</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>Complex</td>
<td>4</td>
<td>9 (11)</td>
<td>2 (11)</td>
<td>2 (17)</td>
<td>21</td>
<td>Yes</td>
</tr>
<tr>
<td>7</td>
<td>Linear</td>
<td>1</td>
<td>7 (8)</td>
<td>2 (11)</td>
<td>2 (17)</td>
<td>28</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>Complex</td>
<td>3</td>
<td>11 (13)</td>
<td>1 (8)</td>
<td>1 (8)</td>
<td>16</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>Linear</td>
<td>1</td>
<td>5 (6)</td>
<td>1 (5)</td>
<td>1 (8)</td>
<td>96</td>
<td>Yes</td>
</tr>
<tr>
<td>10</td>
<td>Linear</td>
<td>2</td>
<td>5 (6)</td>
<td>1 (5)</td>
<td>1 (8)</td>
<td>10</td>
<td>Yes</td>
</tr>
</tbody>
</table>

*aApp store could be either the Google Play Store (2, 4, 6, 9, and 10) or the Apple App Store (1, 3, 5, 7, and 8).*
Start of Search

The start of the search differed among participants in 2 ways. First, the participants used different devices for the search. Of the 10 participants, 1 started the search on a laptop (switching to a smartphone later on), 2 used a tablet, and the remaining participants searched for an app on a smartphone. Participants starting the search on a laptop or tablet indicated that they thought a bigger screen was somewhat easier for searching and reading.

The second difference was related to the place where the participants started their search. Most participants (8/10) went straight to the app store, whereas some (2/10) participants started their journey in a web browser, using a search engine to visit ≥1 website to gain information about apps for smoking cessation before going to an app store:

Because I don’t know what I’m looking for, it’s nice to spend some time online reading. That gives me some “language”, some inkling of what to think of, and after that, I go on to the list of apps. [participant 6]

Search Field

Every participant began by using a search function (either an app store search field or a search engine such as Google). Of the 10 participants, 9 started the search with a Dutch query, and 1 initially used English terms (Multimedia Appendix 5). Seven participants clicked on a query offered by the autosuggestion. Four participants returned to the search function later in the process to use another query (using different terms or switching to English or Dutch) in an attempt to filter the search results (eg, on free or skin) or to search directly for a specific app (by name).

Search Results

After entering a search query, users received a number of results. All participants who searched the Google Play Store and those who searched the Apple App Store with an English search term received only smoking cessation apps in the results. However, Apple users who used a Dutch query received a mix of smoking cessation apps and other unrelated apps, such as Stop Motion apps. One of the participants scrolled down to the 32nd app in the search results list; of the 32 apps, 17 (53%) were not smoking cessation apps.

The information per app in the list of search results in the Google Play Store was limited to a logo, title, rating (number of stars), price (if applicable), and developer name. The Apple App Store additionally provided screenshots but omitted the developer’s name. Some participants (4/10) indicated that they thought the information was scarce. For example, they stated that all the ratings were basically the same, and thus unhelpful, and that the other information was hardly useful for making a proper choice. In addition, some remarked that a small text about the functionality of the apps and a means of filtering the results on price were lacking.

The decision people made in this process step was to click or skip an app in the search results list. Most participants (8/10) mentioned that they based their decision on ≥1 available information cue. People most commonly used screenshots (only in the Apple App Store), ratings, price, and the name of apps; however, click-or-skip decisions were also based on logos and developer names. Some participants (2/10) systematically opened app detail pages by first clicking the first app, then the second app, and so on. Hence, these participants did not use any information cues for their click-or-skip decisions.

For the participants who explicitly mentioned using the information cues, we discerned 3 main reasons for clicking or skipping apps. The first was a positive or negative evaluation of some aspects of the app that was reflected directly in the information cues. For example, this was an evaluation of the design of the app based on screenshots, the popularity of the app based on rating, or the trustworthiness of the developer based on developer name (ie, Tribemos Institute). Sometimes, screenshots and app names provided some information about the functionality of an app, on which people based a click-or-skip decision. For example, the term audioBook in app names could attract or put off participants. Overall, people clicked an app when the evaluation was positive (design attractive, developer a trustworthy party, rating high, and desirable functionality) and skipped apps when the evaluation was negative (app costs money, design unattractive, rating low compared with other apps, and functionality undesired).

The second reason to click on apps was to check something. For example, several participants clicked on apps to check whether the app was, in fact, a smoking cessation app. In addition, half of the participants indicated some confusion over whether they had already opened detailed information for certain apps at some point during the search. To check this, they clicked or reclicked an app in the search results. Furthermore, one of the participants clicked apps out of curiosity and a wish to check what an app was about (triggered by such things as hypnosis in the app name, a combination of a trustworthy source and low rating, or a funny name or concept). Finally, one of the participants clicked some apps because of a personal conviction that “one has to check something out to judge it” (participant 10), although the information in the search results did not trigger a particularly positive evaluation of the app.

The third reason for clicking or skipping an app was based on the participant’s imagined idea about the working of the app. On the basis of the available information, some participants immediately formed a picture of how the app would work and subsequently clicked on the apps they thought were right for them and skipped the apps they evaluated negatively. In some cases, this interpretation of information led to a decision to skip based on nothing more than a logo, app, or developer name:

The little man here, this one, the green one, kicking his cigarette...I would click that one sooner than this woman with “Quit Buddy”. […] She’s going to ask you nicely all the time, I think, or in any case, [she is going to tell you] “well done” all the time. All the time these motivational things. I couldn’t take that very well, I think. But that’s my first insight. Yeah, I don’t know, that [other] one kicks your ass, I guess. [participant 4]

https://humanfactors.jmir.org/2022/2/e32628

JMIR Hum Factors 2022 | vol. 9 | iss. 2 | e32628 | p.38

(page number not for citation purposes)
were associated with the free apps. These participants costs of a free app in particular. Another important information need was negative about the apps in general and about certain functions well technically and whether other users were positive or most participants tried to determine whether apps functioned counters; challenges; badges; and chat functions. Furthermore, specific functions of apps, such as time, cigarette, and money and how they work by focusing on information about the foremost, all participants paid attention to what these apps do apps in general and of specific apps in particular. First and cessation apps exploring the information about the functioning of smoking and continue the search. We describe each action in more detail in practice, and (5) everyone eventually decided to either imagined what using certain functions would be like for them functions and characteristics, (4) some participants also used both. Most of the time, participants browsed the information; however, sometimes, they went in active search of particular information about, for example, costs or user-friendliness. While going through the detailed app information screens, participants performed several actions. The most important actions were as follows: (1) they explored information about the functioning of smoking cessation apps, (2) some participants tried to assess the trustworthiness and personal relevance of the information itself, (3) participants formed opinions about diverse functions and characteristics, (4) some participants also imagined what using certain functions would be like for them in practice, and (5) everyone eventually decided to either download an app or leave the detailed app information screen and continue the search. We describe each action in more detail in the following sections.

The primary action on the detailed app information screens was exploring the information about the functioning of smoking cessation apps to create a mental image of smoking cessation apps in general and of specific apps in particular. First and foremost, all participants paid attention to what these apps do and how they work by focusing on information about the specific functions of apps, such as time, cigarette, and money counters; challenges; badges; and chat functions. Furthermore, most participants tried to determine whether apps functioned well technically and whether other users were positive or negative about the apps in general and about certain functions in particular. Another important information need was the price of a free app. Many participants wanted to know what hidden costs were associated with the free apps. These participants were looking for information about the difference between free and paid versions of the same app; whether one had to start paying over time; and whether paying for an upgrade would get them extra functionality, quality, or just the elimination of annoying advertisements and pop-ups offering upgrades. Finally, several participants looked for information about the quality and professionalism of apps to estimate their trustworthiness. Cues for a trustworthy app could be the name of the developer (known institutions and familiar names generally inspired trust), beautiful design of the app, mention of a scientific foundation, or reactions by the developer to reviews:

There’s always a reaction [from developer to reviews] too, right. They always give a...That’s definitely positive. Professional. Like, at least he’s involved in his own app and taking it seriously. [participant 1]

For some participants, a second action while examining the detailed app information screens was trying to assess the reliability of information itself. Half of the participants were engaged in estimating reliability to some degree, which was particularly complicated for reviews:

But then again, I don’t really know how that works [...] actually, with apps and with reviews. [...] Yes, [I don’t find it credible] that there are so many. [...] I don’t really believe that all of that is true, what it says there. Of course, it’s also just that it could be someone from Vietnam, who gets paid to write reviews there. I think so. Or, I don’t know from which country... [participant 4]

Furthermore, at some point, half of the participants tried to estimate whether the information was relevant to them in their search for an app:

“I made a back-up and put it back.” [...] Oh, that’s just someone who doesn’t know how to [...] transfer that to their new phone...That is not applicable to me. She was actually more critical of her inability to install a new phone than of the app itself. [...] That’s not a review of the app. Yeah, so then I think, yes, I can sit and read all that nonsense, but it comes down to how it ultimately feels and pleases me in terms of use. [participant 10]

Third, while examining the detailed app information screens, all participants formulated opinions about functions and characteristics. These opinions varied from person to person. There was consensus on some functions: the counters and badges were positively regarded by many participants, and the inability to choose one’s own quitting date, even if it were in the future, was regarded negatively by most participants. Opinions varied greatly regarding some functions or characteristics:

I’m more one for shock therapy, like: “Stop, stop now! You’re getting cancer!”, like that. [...] seeing a rotten toe, or something, you know, getting eye cancer from it, that sort of stuff. That impresses me, you know [...] So you have to motivate me, or yeah, punish me, motivate me with my health. [participant 9]

Yeah, you know what the crazy thing is? Yeah, that sounds terrible, I don’t know if you’ve heard it before,
but you know the heart attacks and the lungs, yeah, that doesn’t motivate me. Is that bad? [...] This is a very threatening one, with the number of deaths since you stopped smoking and... But that’s not my motivation. [...] they’ve gone out of their way here to make you very afraid in any case, but that doesn’t work for me. [participant 5]

Paying for apps was a topic on which all but one of the participants gave their opinion. Of the 10 participants, 4 (40%) were prepared to pay for an app but only if it bought them the extra functionality they desired or if it was a guarantee for a high-quality app; 3 (30%) were not strictly unwilling to pay for an app but thought the free ones would do just fine; and for 2 (20%) participants, paying for smoking cessation apps was an absolute no go.

Fourth, next to exploring functions, assessing the reliability and relevance of information, and forming opinions, a number of participants imagined what using certain functions would be like in practice. They tried to imagine how and in which situation they would use a specific function:

“Track your cravings and learn how they can get better over time”. So apparently, I can register when I’m craving a cigarette. That’s kind of interesting because then I can measure it for myself... I know where my weaknesses lie, but [...] I find it interesting because I do think it is fun to do self-examination [...] I do think it’s a nice feature, but... I don’t think I will make very active use of it, if it’s, like, half past one in the morning and I think “I feel like having a cigarette”, I don’t think I will grab my phone and think, “Half past one in the morning, I’m sitting here on a terrace [...] a glass of wine in my hands and I feel like having a cigarette now. Ohh...” I don’t think I’m going to do that. [participant 8]

Finally, at some point in the search, every participant had to decide to either leave the detailed app information screen or download the app. This choice was the result of the four abovementioned actions: exploring and imagining resulted in a mental image of smoking cessation apps; opinions about the functions, characteristics, and trustworthiness of the apps; and an assessment of the reliability and personal relevance of information, resulting, in turn, in decisions to either download the app or leave the screen.

In total, the 10 participants opened and left 85 information screens (range 2-25). In some cases, the reason for leaving a detailed app information screen would be practical, such as wanting to see more apps, comparing some apps with others, or an app turning out not to be a smoking cessation app. However, most of the time, people left these screens as the assessment of (some aspects of) the app came out negative. The most common reason for appraising an app negatively was finding a particular function or feature in the app unappealing, unhelpful, or not in accordance with (developing) wishes or requirements. In addition, doubts about the reliability of the app or a certain approach played a role in the negative assessment. Furthermore, bad reviews from others or a small number of reviews and ratings often caused participants to assess an app negatively and leave the detailed app information screen. For half of the participants, the presentation of information in itself played a role at some point. For the participants, language and spelling errors, poor (automated) translations, and a perceived cluttered structure of text or screenshots were the reasons for leaving a detailed app information screen and continuing the search.

At some point, every participant chose to download an app. The first app was downloaded after participants had viewed, on average, 5 detailed app information screens (range 1-9). Overall, 4 participants downloaded ≥1 app (Table 1). Overall, 2 of them (participants 1 and 8) downloaded multiple apps (3 and 6 apps, respectively) to find a specific desired requirement. One of the participants (participant 7) downloaded 2 apps wanting to view them both live and then decide which one to keep. One of the participants (participant 6) downloaded an app that he wanted to listen to just for fun in preparation in addition to the one he planned to use during the quit attempt. Most participants indicated that they were downloading apps as part of the search process to explore the apps to see what they were like in practice.

The App

All participants opened the apps that they downloaded. Two participants (participants 3 and 6) decided, immediately after opening them, not to explore the chosen apps on the spot. One of them wanted to enter data privately after the interview, and the other did not want to start the trial period at that particular moment. The remaining 8 of participants explored their downloads.

All participants started exploring by clicking on the menu options and buttons to see (and discover) what the app did, how it worked, and what the possibilities were. Exploring the apps resembled the exploration of the information on the detailed app information screens, in the sense that participants stated what they liked or not and what they thought would be helpful. Similarly, the participants imagined whether and how they could potentially use certain functions in practice. In addition to exploring, a number of participants actively sought the functions or features they desired.

Exploring the first download led to the decision to either keep or discard the app. Of the 8 participants, 5 remained (or became more) enthusiastic about their first download after exploration and chose to keep the apps (participants 2, 4, 7, 9, and 10). One participant (participant 5) ran into an Upgrade to Premium pop-up, which discouraged proper exploration of the app and made her continue the search without discarding the app. This participant went back to the app store and looked at 17 more detailed app information screens before returning to the initial download and exploring it more thoroughly. After the second exploration, the participant concluded that the app was truly the best one she had encountered and that it actually met her wish or requirement. Of the 8 participants, 2 (participants 1 and 8) discovered something they really disliked about their first app during the exploration and decided to discard the app and continue their search:
I don’t even get to choose tomorrow! Or do I have to...? It says here: “Last year” So I can go into the past, but I MUST stay in the now. [...] I don’t have a choice. I can’t say I want to stop next week because I’m starting medication now for example. [...] They just assume...I want to download the app and they just assume “now you don’t smoke anymore”. Yes, now I’m already inclined to...I’m curious how that works in the other apps. Whether they also just say “bam” [...] Well, what irritates me most, or bothers me, is that I am not allowed to choose when I want to stop. [...] I’m just going to find another one. [participant 1]

From that point on, these 2 participants (participants 1 and 8) changed their way of searching. They had chosen their first downloads as, based on the information they had viewed on the detailed app information screens, they found certain features fun and attractive, could imagine them as helpful, and found the design appealing. After they came across the aspects in their first downloads that were so objectionable (the setting of the quit date in the future and the costs of the app), their search turned into a hunt, really only paying attention to that one requirement. Both decided to keep a downloaded app as soon as they found one that met the requirements.

Choosing to keep an app (and thus stopping the search) was related, in the first place, to satisfaction with certain characteristics and functions but also to a sense of saturation. Half of the participants indicated that they felt they had explored enough apps. For some Apple users, this meant that they felt they had viewed the full range of products as the app store returned a limited number of relevant results for a Dutch search query. For a few participants, saturation occurred as their search had taken quite some time, and they had viewed a lot of information. One of the participants was saturated after reviewing a self-pronounced delimited set of the first 10 apps in the search results list.

For a number of participants, in addition to satisfaction with the functions and features and saturation, feeling certain emotions played a role in choosing and discarding an app. Several participants were simply excited enough about the app they had downloaded, opened, and explored to stop searching. One of the participants was surprised to have eventually found exactly what she was looking for. Two participants were tired of searching; 2 others were extremely frustrated during the search and were so relieved when they had finally found something that met their needs that they immediately ended the search:

Whaa! Help. What frustrations...My god. [...] Uhm so no, now I’m like...[...] But what I’ll try one more time is to enter “quit smoking” now instead of...See if I get completely different results now. [...] We’ve already seen this one, we’ve also seen that one, we’ve also seen that one...Not this one. [...] I’m not seeing anything annoying yet, so. I have my health things, I have my milestones. And apparently this is free so then...great. Okay, well, we have an app. And I don’t want to think about it any further now [laughs]. [participant 8]

End of Search

Eventually, every participant ended the search with at least one app and the intention to use it during the next cessation attempt. More than half of the participants felt that they could not still properly judge the app and its usefulness before using it for some time. Several participants indicated that if through use, they would discover that they did not like the chosen app after all, they would have no problem getting rid of the app, switching to another app, or starting to look for other support tools (such as medication or e-cigarettes) for the cessation attempt. This low threshold for discarding the app seemed to be related to the apps being free.

Looking over the process as a whole, across the separate process steps, we observed additional factors that played a role in the choices people made, such as ranking and rating, feelings, and errors in information processing. We describe each factor in more detail in the following sections.

First, the roles of both ranking and rating in making choices were somewhat ambiguous. Apart from one participant, none literally named ranking as important information in their search. Moreover, half of the participants scrolled down further than rank 20 in the search results, and approximately a quarter of the viewed detailed app information screens were those of apps with a ranking >20 (maximum 94). Thus, during the search process, ranking did not seem to play a role for our participants. However, the apps that the participants ultimately chose to use were all in the top 10 in terms of ranking; therefore, ranking did seem to be of influence on the outcome. Similarly, for rating, although many participants also viewed information screens of apps with very low ratings (range 2.3-5) and of apps with no rating (because of too few reviews), for some participants, we observed that rating played an important role in the choice of clicking or skipping apps in the search results overview. Moreover, the average rating of the chosen apps was 4.5 (range 3.9-4.8) stars, whereas the average rating of all viewed apps (that had ratings) was 4.3. Once participants arrived on the detailed app information screens, rating seemed less important for some as their focus was drawn to functionalities, design, or other features that excited them or that they considered important.

Second, in addition to rational arguments for choosing to click or skip, leave a detailed information screen, or download or discard an app, almost every participant indicated somewhere in the process that they made a certain decision as something did or did not feel right. For example, one of the participants did not have a good feeling about a particular app while reading the information in the search results and on the detailed information screen. He associated the developer’s name with a treatment for alcohol addiction, the app came across as American (“not my favorite...um, people, in terms of attitude and behavior and such” [participant 10]), and he found the use of the word PhD in the developer’s name annoying, as well as the mix of Dutch and English in the description. Strictly speaking, none of these things had anything to do with the...
content or quality of the app; however, nonetheless, they discouraged him from choosing the app.

Finally, we observed the influence of errors in information and information processing on the decisions participants made throughout the process. For all participants, somewhere in the search process, something went wrong. It could be that people missed something in the information, did not read it properly, misinterpreted it, or misremembered it. In addition, the information itself was sometimes unclear, incomplete, or hard to find. As a result, people occasionally drew wrong conclusions and made wrong assumptions. A number of times, we observed that choices (click, skip, download, or discard) were based on a judgment that was formed on information that was misread, misinterpreted, misunderstood, or misremembered or as information could not be found.

In many cases, these kinds of decisions did not necessarily have any kind of impact. For example, one of the participants (participant 3) mixed up all kinds of information she had seen and read. After making the choice, she mentioned that she thought usability was important, as well as the large number of reviews (as to her, that was an indication of many downloads and, thus, popularity, which she considered important). She remembered reading in the reviews of the app she chose that the app was user-friendly. However, the recorded images showed that none of the reviews said anything about user-friendliness. She also remembered that one of the apps she had not chosen had very few reviews. However, the images showed that, of the 6 apps this participant reviewed, the one she referred to was one of the apps that had the most reviews, and the app she had chosen turned out to be one of the apps with the fewest reviews. Thus, it seemed that this participant had misremembered that negative features belonged to apps she did not choose, and features she found positive belonged to her chosen app.

In some cases, errors in information (processing) led to a profoundly negative experience or an inferior choice of app. For instance, one of the participants (participant 8) had a very frustrating search caused by not reading carefully and as certain information was hard to find. She mistakenly wrote off several apps that were fully compliant with her requirement for a free app with certain basic functionality. Another participant (participant 10) who did not have a good feeling about a particular app wrongfully assumed the things that made him feel bad about the app (the app was not American but British, eg, and the mix of Dutch and English in the description was caused by an app store functionality and not chosen by the developer). If the participant had not made these errors in information processing and had not written off the app for these reasons, he would have had a higher quality app in this one than the one he ultimately chose.

Thus, over the whole process of ranking and rating, feelings and errors in information processing had some influence on the choices people made. On the other hand, we observed that privacy-related information was not important for any of the participants anywhere in the process. None of the participants viewed the information about permissions on the detailed app information screens. After opening the downloaded apps, almost all participants instantly agreed with their privacy policies, terms, and conditions. A total of 2 participants first quickly scrolled through the text before giving consent but also immediately indicated the futility of that action:

Yes, actually I always just “agree” [laughs]. I don’t quite feel like reading all the way through, that ehh. [...] I just think, it’ll be fine. [participant 4]

I did read for a while, but then I couldn’t choose anything there. I mean, that was it, so yes, I couldn’t do anything else there except click on it because otherwise I couldn’t continue. [participant 9]

After 2 Weeks

After 2 weeks from choosing an app, of the 10 participants, 6 (60%) had used the app to some extent, of whom 4 (67%) had also quit smoking (Figure 2). Alternatively, one of the participants had quit smoking without using the app. Finally, 3 participants had not used the app and had not quit. Of these 3 participants, 2 had already indicated at the end of the interview that, because of personal circumstances, they were not confident that they would actually start their quit attempt right after the interview (both scored 1 on the confidence ruler), and 1 had not managed to start the quitting attempt, although she was rather enthusiastic about the app and had been moderately motivated to quit during the interview (score of 5 on the confidence ruler).

The 6 participants who had actually used their app had enjoyed occasionally using some functions in the app (the distraction game and the motivation cards) or viewing certain information (the counters and health information). Of the 6 participants, 5 had only used a small number of functions and to a limited extent, and 3 of them indicated that the app could not do much other than count days, cigarettes, and money; however, these participants also immediately admitted that they had not actually explored the app thoroughly. They realized that there might be more functionality available in the apps. For these participants, the app had not played an important role in quitting. However, of the 6 participants, 1 had used the app more extensively and indicated that the app had supported him in his quitting attempt.

In retrospect, what the participants remembered most about finding an app for smoking cessation was that many apps, more or less, offered the same functions and looked similar, making it hard to distinguish among them. We also saw this at times during the search when participants tried to remember the features of a particular app. At such times, it appeared that people mixed up (information about) apps and, in some cases, did not remember whether the app had already been viewed or even downloaded. Combined with hard to find, limited, or absent information, sometimes, the only way to find out about something was to download the app. Consequently, a number of participants indicated that they thought it actually takes (too) much time, effort, and energy (in some cases because of negative emotions) to really look for an app properly. For some, the frustration of the search was still fresh in their minds:

Going back to look for another app? No, no, I found that process so tedious, already after 5 minutes. I’m really not going to do that again, no. Haha, no, I found searching for those apps, oh my god...
The participants who had used the app intended to leave it on their phones for now, mainly for the counters. Participants who had not yet used their apps intended to save them for their next quit attempt. Thus, no one expressed any intention of going back to the app store or looking for another app.

Figure 2. App use and quitting success after 2 weeks.

<table>
<thead>
<tr>
<th>Used app</th>
<th>Did not use app</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quit smoking</td>
<td>Did not quit smoking</td>
</tr>
<tr>
<td>Participants 3, 4, 9, 10</td>
<td>Participants 2, 8</td>
</tr>
<tr>
<td>Participant 7</td>
<td>Participants 1, 5, 6</td>
</tr>
</tbody>
</table>

### Transformations

From the start of the search until 2 weeks after searching and selecting and, in some cases, using the app, we observed transformations in 3 distinct areas. Over time, participants gained knowledge of the workings of smoking cessation apps and simultaneously developed clearer ideas about their personal wishes and requirements for an app. However, confidence and trust in the ability of these apps to really help while quitting remained quite low or even decreased. We describe the changes in each area in more detail in the following sections.

Before starting the search, every participant could think of at least one or two basic functions (eg, counters and notifications), remembering these from earlier experiences with smoking cessation apps or from stories they had heard from other people. However, none of the participants, including those who used an app in the past, had any knowledge of what currently available smoking cessation apps were able to do and offer. During the search, every participant recognized the basic functions and also discovered new functions and features in smoking cessation apps they had not known or realized existed beforehand. After the search, all participants felt they had a more complete picture of the range of smoking cessation apps, what they can do, and what the landscape looks like.

For all participants, learning about functions and features went hand in hand with forming ideas about what they wanted and did not want from an app. While gaining knowledge, participants developed ideas about what they would like, enjoy, or (on the contrary) find irritating and annoying about an app (wishes), as well as what they thought would or would not help them and, thus, be important in an app (requirements).

The development of wishes and requirements could even continue after choosing an app. The functions and features participants had liked during the 2 weeks of using the apps were often things they had already noticed during the search. However, in some cases, participants were surprised by the fun aspects that they had not seen information about while searching. Notably, a few participants were surprised to find certain functions in the app that motivated them and changed their minds about those functions. For example, one of the participants gave his opinion about a specific feature while exploring the app during a contextual inquiry:

> Well, the way that works, I just find that weird. Because [...] if you accidentally shake your phone, another one of those things will appear. I don’t need that. [participant 10]

After 2 weeks, the same participant said the following:

> If you have your phone in your hand and you shake it too hard, then it automatically gives those quotes and stuff on the screen, so to speak. Sometimes when you’re not even [...] engaged with it, and you pick up your phone, then suddenly there’s this thing on the screen, so to speak [...] I think that’s a good thing. You don’t really get the chance to forget about it, or to let your attention wane, so to speak. So that, yes, for me that works. [participant 10]

The transformation of confidence in the helpfulness of smoking cessation apps was slightly more fuzzy, with no clear patterns or groups (for an impression of the fuzziness in the changes in confidence, see Figure 3). Generally, confidence was not very high for any of the participants beforehand. For some participants, this was because of a mediocre experience with these types of apps in the past. For almost all participants, low expectations about the ability of apps to help with quitting seemed linked to low confidence in the ability of cessation aids to help during a quitting attempt in general:

> Well, it’s certainly not going to be the ultimate remedy. I’m too stubborn for that anyway and I know that I, I have to do it myself. And aside from someone coming and sitting next to me all day and knocking every cigarette out of my hands...There’s no way the app is going to do that. [participant 10]

Immediately after choosing an app, participants were asked to estimate their confidence that the chosen app would actually help them quit smoking. For some, confidence had increased slightly compared with the confidence participants indicated having in smoking cessation apps in general before the search; for a few, it was similar; and for one, it had dropped significantly. Although this participant felt that he had chosen...
the best app available, he had become disappointed in the landscape of these types of apps through the extensive search:

...especially telling that during my search, no really serious things come up. I think that’s a very simple fact that says a lot. For example, that there doesn’t seem to be an app that costs 100 euros a year. That makes the domain serious, that world, that makes that there is a landscape. That there are things that cost 100 euros, things that are free or a few euros. Then there would be something of a landscape, and now there is not. Actually, we have seen 2½ ways of...That is, a very simple counter and an app that has something of interaction in terms of cravings. [...] Yeah, the disappointment that I feel now at the end, like: “yeah, it’s just not there, or something, that [serious] app.” [participant 6]

For a couple of participants, the disappointment in the chosen app manifested only after 2 weeks. During the extensive search, they felt that they had looked at enough apps (sometimes at everything there is) and had chosen the best app from the range on offer, only to discover during use that even the best was not very good:

I did hope [that this app could hold my interest]. I think, you know, I was going to do a really good search, and that’s what I did with you at the time, but no, [...] no. [...] There’s nothing innovative in it. [...] Maybe I thought, “well, this is it then” because I chose very consciously [...] and didn’t simply take the first one I could find. Then I think, well, this is going [...] to be the Columbus’ egg. But it turned out not to be. [participant 9]

A few participants, who had not had a high opinion of quit-smoking apps 2 weeks earlier, found their lack of confidence confirmed during use:

I don’t find the app reliable, because after every day it says: “you will live 60 minutes longer”. Then I think: “yeah, bullshit probably”. Or just things where you think: yeah, I don’t know...this is probably just not true. Anyway, it’s kind of funny to see [...] [but] I don’t take it very seriously. [participant 8]

For most participants, even for those who were enthusiastic about certain aspects of the app, confidence in the app’s ability to help them quit smoking did not increase after 2 weeks compared with the confidence they had immediately after searching. Again, most participants indicated that quitting smoking is simply hard and a matter of perseverance and discipline and that no app in the world can do anything to make it easier:

I do adjust the grade down a bit, to a six [instead of an eight] in the sense that it really helps to stop smoking. [...] you can’t stop smoking just by using an app. I mean, there’s more to it. But there’s nothing the app can do about that. [participant 2]

Figure 3. Confidence in the helpfulness of smoking cessation apps per participant at 3 points in the search process.

Discussion

Principal Findings

This study set out to explore the process of searching and selecting apps for smoking cessation and map the range of actions and the reasons for those actions during the search, focusing on both the information needs and experiences of those who aim to find an app. The empirical findings in this study have expanded our knowledge of the process, information needs, information processing and decision-making, and transformations that occur when searching and selecting apps for smoking cessation.

With regard to the process, we found that participants thoroughly searched for an app that they expected to contribute to smoking cessation. All participants were actively involved in exploring, evaluating, imagining, comparing, searching,
assessing, choosing, and navigating. The comprehensiveness of the search was reflected in several aspects. Many participants continued to look at app information screens and download apps to find something they were somewhat confident in, even if they were fed up or frustrated. The most extensive searches involved using multiple search terms and going back to previously viewed app information screens to discover more and compare apps. Participants viewed many detailed app information screens and scrolled far down the list of search results. No one used a Take the First heuristic [30]; one of the participants chose an app after viewing 2 detailed app information screens; however, otherwise, everyone viewed ≥5 screens. Many participants read texts thoroughly. Only 1 participant hardly read at all and chose to download apps based on heuristic cues such as ratings and pictures. Most participants also explored the downloaded app as part of the search process. Searches took quite some energy: although there was laughter and participants were generally happy to be looking for an app, for some, the whole process caused negative emotions such as frustration, irritation, and disappointment. Furthermore, several participants indicated fatigue at the end of the search. Overall, it appeared that although confidence in the helpfulness of smoking cessation apps was low, everyone made a real effort to find the best possible app.

The search process of our participants was far more extensive than we had expected based on one of the few studies on choosing apps [30]. In that study, only 16% of the participants used a strategy of viewing >1 detailed app information screen before making a choice when choosing (among others) a running app. This former study was conducted in a laboratory setting, used special research devices, and was conducted with participants who did not necessarily have use for a running app. Our participants may have been more invested as they were looking for an app they actually intended to use on their own devices and in their own personal context. It was also notable that all 10 of our respondents chose and downloaded an app with the intent of using it, whereas uptake was found to be far lower in other studies [35]. This result may also be related to the level of investment. Alternatively, it may have been the formulation of the task (to search for an app like you would do if I were not present). Although we took care to tell the participants that deciding not to download an app was also an option, emphasizing that choosing an app was certainly not required to end the task, the task may have been leading. Another result we did not expect was the extent to which participants scrolled down the list of search results. It is well-known from research on internet searches that people never scroll down further than the third page [53,54]. However, what is consistent with research in this field is that the first and second results were viewed most often.

Second, with regard to information needs, our findings show that participants mainly paid attention to and went in search of information about the functioning of smoking cessation apps. In doing so, they mostly paid attention to what these apps do and how they work, whether apps functioned well technically, whether other users were positive or negative about the apps and their functions, the price of a free app, and quality and professionalism of apps. In addition, some participants tried to assess the trustworthiness and personal relevance of information itself.

Information about the functionalities included in an app (eg, counters, community, Facebook, and coaching) and what features the app has (ie, design and price) was easily found by most participants and could be obtained from descriptions, screenshots, and reviews and by exploring downloaded apps. Information about the content and technical quality of apps could not be gleaned directly from descriptions and screenshots and was, therefore, more difficult to find. Some participants dug deep to assess whether app developers had proper expertise, whether the intervention was good and reliable, and whether it was based on a scientific foundation but often could not find any information about it despite the extensive search. Information about the true costs of free apps was equally hard to find.

Nearly everything we have observed in terms of information needs is consistent with previous research. As in previous research, our participants paid attention to features such as monitoring, feedback, goal setting, rewards, reminders and prompts, progress sharing on social media, coping games, health and statistical information, communication style, and ease of use [31-33]. Recently, a study by Szinay et al [35] showed that people also primarily pay attention to these potentially engaging characteristics when searching for health apps. In addition, similar to participants in other studies, our participants also paid attention to immediate look and feel, design, other people’s star ratings or reviews of apps (social proof), and costs during the search [31,35]. However, the considerable focus on the hidden costs of free apps (eg, whether paying for an upgrade would get you extra functionality, quality, or just the elimination of annoying advertisements and pop-ups offering upgrades) is something we have not seen in other studies. This insight is an addition to the factors that people consider important during the uptake of apps for smoking cessation.

Furthermore, this study brings nuances to existing insights from the literature regarding the importance of ranking and rating. Previous research [28,29] has shown that apps with (among other things) high rankings and high ratings are downloaded most often. Although the apps chosen in this study all have high ratings and rankings, this was not what participants paid the most attention to during their search. Participants mainly wanted information about the features of the apps that they expected to be fun or helpful. Individuals looked for specific characteristics of an app (eg, functionality, appearance, and price) and simply started at the top of the search results. Therefore, although ranking was seen by a few participants as a useful source of information for selecting apps, the influence of ranking is clearly noticeable, as starting a search at the top of the list is simply convenient and obvious. This leads to the fact that in the Google Play Store, for example, a few dozen apps at the top of the ranking account for almost half of all downloads [54]. More than 85% of all health apps are found much less often, rarely, or never [55]. This is potentially a loss, as that 85% of apps may contain exactly the functionalities and features that someone is looking for.
Third, in our results regarding information processing and decision-making, we observed that participants had to make several decisions during the entire process. In addition to smaller ones, such as choosing search terms, every participant chose to click on or skip apps in the list of search results; leave a detailed information screen or download an app; and finally, after downloading, keep or discard it. To make these decisions, participants needed to understand, interpret, and remember information, form a mental picture of smoking cessation apps, and continually adjust that picture based on new information. Furthermore, choosing an app involved thinking about wishes and requirements and formulating opinions about the functions and features of apps. Some participants also imagined what using certain functions would be like for them in practice. Overall, this seemed to be quite a cognitive load, as without the participants realizing it, they often made mistakes in information processing. A number of times, we observed that choices (click, skip, download, or discard) were based on a judgment formed on information that was misread, misinterpreted, misunderstood, or misremembered, ultimately affecting the final choice for an app. These findings are in line with the known deficiencies in human thinking and decision-making [56,57], including, among others, restricted capacity and forgetting [58]. To the best of our knowledge, the influence of errors in information processing on choosing health apps has previously not been explored and recognized in other studies in this field.

Finally, during searching and selecting, we observed transformations in the areas of knowledge, wishes and requirements, and confidence in apps. Knowledge increased from knowing 1 or 2 basic functions before starting the search to participants feeling they had a full picture of what smoking cessation apps can do and offer. While gaining knowledge, participants developed ideas about wishes (likes or dislikes) and requirements, which were eventually important in deciding which apps to download, discard, or keep. Notwithstanding this development of knowledge, wishes, and requirements, confidence in smoking cessation apps did not vary much if we compare the participants’ estimations before, immediately after, and 2 weeks after the search. For some, confidence was slightly higher immediately after the search, leaving participants rather optimistic. However, that rise was nullified after the 2 weeks of use, with confidence returning to the level before the search or even lower. For some participants, confidence in smoking cessation apps as useful aids had already decreased immediately after the search.

Although it is fully expected that people go through a transformation in knowledge, wishes, and requirements during the decision process [59], to the best of our knowledge, this has not been reported before as an essential part of the search for health apps. We reflect on this in the Suggestions for Further Research section. With regard to trust in smoking cessation apps, we confirmed what the study by Regmi et al [60] put forward as a potential threat to smoking cessation apps. In an analysis of strengths, weaknesses, opportunities, and threats, they postulated the loss of trust from users because of the incongruence of perceived app abilities and actual functionalities.

Strengths and Limitations
This study’s additions to the literature are primarily the result of our application of contextual inquiry, a method that is not often used in comparable studies. By using contextual inquiries, we were able to study the act of choosing an app in a situation as naturally as possible. Participants could search on their own devices at a place and time that was most convenient for them. This may have increased the likelihood that participants would feel at ease, be honest and open, and understand and accurately remember information, which, in turn, would contribute to data quality [61]. Furthermore, the mindset created by the contextual inquiry’s specific basic principles of apprenticeship and partnership facilitated curiosity, humility, interest in, and respect for the respondent, which are generally seen as success factors in conducting interviews [62]. Moreover, close collaboration with participants throughout the research process is thought to produce credible data [63].

We extended our contextual inquiries by video recording the screens of the participants’ mobile devices and audio recording their comments. This allowed us to detect that there is often a discrepancy between what people think they see, read, understand, and remember and what actually is on the screens. These double recordings also enabled us to observe that choices are frequently based on the faulty processing of information.

We also consider the inclusion criteria for our participants as a strength of this study. The inclusion of participants in the study who wanted to search for an app to actually stop smoking caused the respondents to be more invested in the task and made the task less artificial. Making observations of actual, realistic behavior in a natural context may have contributed to the ecological validity of the research [64].

Notwithstanding the strengths of the methodology, our chosen approach also has some drawbacks, which create a number of limitations. First, it has been hypothesized that a good rapport between the interviewer and respondent also has downsides and could result in response bias as it causes respondents to ingratiating themselves with interviewers [61]. This could explain why some of our participants indicated that the search during the contextual inquiry was somewhat different from how they would normally search for an app. At the end of the search, 3 participants (participants 3, 4, and 9) indicated that they had chosen more consciously than they normally would have as they had to state aloud why they made certain choices. Three participants (participants 4, 5, and 9) indicated that they had searched a bit more extensively and thoroughly. For 3 participants (participants 2, 7, and 8), the way they had searched for a smoking cessation app was completely different, as normally they would never browse but just go to the app store for a direct search based on an app name. Lastly, 3 participants (participants 1, 6, and 10) searched and found their app in much the same way they would normally do.

In addition, during analyses of the data, we reached saturation [65] when we got to the point where further data collection would not necessarily add anything to the overall story [66]. Even before the analysis of the tenth and last inquiry, we found no new variants of expressions in behavior within the themes or subthemes. The decision to stop further recruitment was
reinforced by the consideration of time investment in recruitment and the chosen methodology. Nevertheless, it is conceivable that studying more people could lead to an even richer description of the search process, people’s actions, and their motivations. For example, one of the participants was diagnosed with Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS) and surprised us by looking at the information in a completely different way than the other participants.

A further limitation concerns the composition of our sample. As the aim of the study was to better understand the process of choosing a smoking cessation app, recruitment was not about getting a representative sample but about composing a group of people in such a way that we could gain different perspectives. By purposive sampling on factors that could theoretically influence the app choice, such as gender [67-69], age [70], and education level [71], we tried to do just that. However, it is thinkable that different cultural backgrounds or other personal characteristics could provide different, new insights.

**Suggestions for Further Research**

This study raises several new questions. During the search, participants gained knowledge about smoking cessation apps and developed wishes and requirements. This finding implies that the search process in itself plays a role in the uptake of apps. This raises entirely new questions about the influence of these transformations on the outcome of the search process, selecting an app: how do gaining knowledge and developing wishes and requirements shape the decisions people make, is it an important part of the decision process, does it lead to different outcomes than a search in which no transformations would take place, do these transformations also occur in less active and thorough searches, and what underlying mechanisms are at play? For instance, as all our participants chose an app with the intention of using it, an active and thorough search may have contributed to more satisfaction with the choice [72] and lower uncertainty and thus have increased the intention of using the app [73].

Another potentially interesting question is one regarding the effect of the number of presented search results. The Apple users who used a Dutch search term were presented with more information screens than Android users (mean 11.6 vs mean 5.2) and downloaded more apps (mean 2.6 vs mean 1.2). A number of participants indicated that they liked the fact that there were not so many results but were concurrently puzzled by the limited results and presentation of irrelevant apps. Experimental research with more respondents might explore differences in experiences, feelings, considerations, and decisions among various numbers of search results.

Finally, the matter of initial use is intriguing. Much research in the field has focused on understanding the factors that influence uptake, such as what people find engaging. The goal of many of these studies is to increase uptake by helping users to, for example, obtain information about things that are potentially engaging [32,33,74,75]. In this study, we saw that participants searched for the functions and features they liked or found useful and that uptake in the sense of downloads was high—every participant ended the search with an app and the intention to use it. However, after 2 weeks, we saw that some of the participants had not even opened the app. Despite successful uptake based on expected engaging functions, initial use was thus not guaranteed, let alone actual engagement and continued use. We suggest that it may be worthwhile to investigate what happens between uptake and initial use. It could be useful if further research takes into account the extra step of initial use between uptake and continued use.

**Implications for Practice**

The results of this study indicate the need to work on the forms of decision support in app stores. We propose a number of suggestions for the design of three obvious solutions to support people in searching and selecting a fitting app for smoking cessation: advanced filters, recommender systems [76], and curated portals [35].

The first solution involves advanced options to filter the search results. In an immense supply, where people want information that is not easy to find, if done properly, filters can make a difference in terms of time, energy, and positive search experience [77]. Choices based on popularity and others’ opinions can be made relatively easily by people themselves. Therefore, filters should focus on the content of apps, taking into account the functions and characteristics of the app. With the help of technologies such as natural language processing [78], text analytics [79], and machine learning [80], it is possible to analyze apps in terms of content and identify the characteristics present in the app. Filters and other tools based on the identified characteristics can easily be included in the user interface of an app store, with terms that are relevant, useful, and recognizable to the user, to help the user choose an app that is valuable to them.

The second solution is recommender systems. In this study, we have seen that participants put much effort into figuring out what functions and features they expect will really help them and that they actually find that very difficult to do. Most participants seemed quite unaware of what they needed to support them in their behavior change. Thus, many choices in our study (click or skip, click or download, or keep or discard) ended up being based purely on a feeling or on what participants found fun, attractive, or funny. However, there can be a discrepancy between what people indicate to prefer and what actually works well for them [81-83]. The possibility of matching apps and participants with a recommender system could theoretically go beyond matching based on what participants like. Recommender systems could be trained with delayed feedback on the effectiveness of the app on the health behavior change, in this case: smoking cessation. Through this training, the system gradually learns which (functions in which) apps work for whom, optimizing the systems’ recommendations on the expected effectiveness of an app, ultimately helping people to find an app that they not only like but that may also work effectively for them.
A solution in the form of curated portals adds value in yet another way when supporting people in choosing an app. First, we have seen that several participants wanted to find a professional, evidence-based app founded on scientific insights. However, information about the quality of apps is almost impossible to find in the app stores. From earlier studies, we know that high-quality apps are scarce to begin with [22,60,84,85] and, therefore, difficult to find in enormous supply. People for whom quality is a criterion would be helped by reliable assistance in choosing. There are reliable sites that users also trust [31,35], such as the GGD App Store in the Netherlands [86]; however, these are not found by users, as this study and previous research have shown [31,35]. An easy-to-find, well-curated site could also help keep people from giving up after a first tried app. It can be a safe and orderly collection where people can return to try a new app if they do not like the first one.

The second argument in favor of curated app portals is data and privacy protection. As in previous studies [35], we also observed that participants hardly glanced at permissions, privacy terms, and conditions. Although people regularly indicate that privacy and data protection are important to them [31,32], in practice, for most, it is not feasible to process and understand this type of information [87]. Even if consumers were to read the incomprehensible terms and conditions, privacy could be incomplete [88]. Moreover, it has been found that many apps, both free and paid versions, display dismal privacy practices [89,90]. As the use of apps depends on the acceptance of the conditions, and many people are not (or cannot be) aware of the risks [87,88], a reliable, independent party that monitors privacy and data conditions is of great importance.

**Conclusions**

The empirical findings in this study add insights into the literature on the process, information needs, information processing, and decision-making and transformations in knowledge, wishes and requirements, and confidence and trust that occur when searching and selecting apps for smoking cessation. Currently, finding an app that contains functions and features you expect to help you quit smoking takes considerable time and energy and can even be a negative experience. At present, app stores do not appear tailored to finding suitable smoking cessation apps, and consequently, people who want to quit smoking need to process a lot of information and make a multitude of choices. In this entire process, errors in information processing creep into and affect decisions. Furthermore, although every participant downloaded an app with the intention of using it (uptake), initial use was lower, and subsequent continued use and engagement were even lower. As such, our findings highlight the need for further research into the factors that affect initial use and into the relationship between active, thorough searches and uptake and initial and continued use. Furthermore, our findings stress the importance of developing helpful tools to guide users through the immense supply of health apps, such as advanced filters, recommender systems, and curated health app portals. Among other things, we suggest the creation of filters and recommendations based on app functionalities and curated portals to guide people to high-quality and trustworthy apps. These solutions could make the search process easier, faster, and more enjoyable for people who wish to find an app that is valuable to them and ultimately effective.

**Acknowledgments**

The authors would like to thank the participants for their time, energy, and candid participation in this study. In addition, they would like to acknowledge the use of DeepL Translator (free version) for translating parts of the text.

**Conflicts of Interest**

None declared.
References


41. Hendriks et al JMIR HUMAN FACTORS 2022 | vol. 9 | iss. 2 | e32628 | p. 50 https://humanfactors.jmir.org/2022/2/e32628 (page number not for citation purposes)
90. Polykalas SE, Prezerakos GN. When the mobile app is free, the product is your personal data. Digit Pol Regul Govern 2019 Mar 08;21(2):89-101. [doi: 10.1044/2018_AJAO-IMAIA-18-0008]

Please cite as:
Hendriks Y, Peek S, Kaptein M, Bongers I
Process and Information Needs When Searching for and Selecting Apps for Smoking Cessation: Qualitative Study Using Contextual Inquiry
JMIR Hum Factors 2022;9(2):e32628
URL: https://humanfactors.jmir.org/2022/2/e32628
doi:10.2196/32628
PMID:
Proposal for Post Hoc Quality Control in Instrumented Motion Analysis Using Markerless Motion Capture: Development and Usability Study

Hanna Marie Röhling1,2,3, MSc; Patrik Althoff1,2,3; Radina Arsenova1,2,3,5, MD; Daniel Drebing1,2,3; Norman Gigengack1,2,3; Anna Chorschew1,2,3; Daniel Kroneberg6, MD; Maria Rönnefarth6,7, MD; Tobias Ellermeier1,2,3,8; Sina Cathérine Rosenkranz9,10, MD; Christoph Heesen9,10, MD; Behnoush Behnia11, MD; Shigeaki Hirano12, MD, PhD; Satoshi Kuwabara12, MD, PhD; Friedemann Paul1,2,3,6,13, MD; Alexander Ulrich Brandt13,14, MD; Tanja Schmitz-Hübsch1,2,3,13, MD

1Experimental and Clinical Research Center, a cooperation between the Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association and the Charité - Universitätsmedizin Berlin, Berlin, Germany
2Experimental and Clinical Research Center, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany
3Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany
4Motognosis GmbH, Berlin, Germany
5Department of Pediatrics, St Joseph Krankenhaus Berlin-Tempelhof, Berlin, Germany
6Department of Neurology, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany
7Clinical Study Center, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany
8Department of Neurology, Vivantes Auguste-Viktoria-Klinikum, Berlin, Germany
9Institute of Neuroimmunology and Multiple Sclerosis, Center for Molecular Neurobiology Hamburg (ZMNH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany
10Department of Neurology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
11Department of Psychiatry, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany
12Department of Neurology, Graduate School of Medicine, Chiba University, Chiba, Japan
13NeuroCure Clinical Research Center, Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Berlin, Germany
14Department of Neurology, University of California, Irvine, CA, United States

Corresponding Author:
Hanna Marie Röhling, MSc
Experimental and Clinical Research Center, a cooperation between the Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association and the Charité - Universitätsmedizin Berlin
Lindenberger Weg 80
Berlin, 13125
Germany
Phone: 49 30 450539718
Fax: 49 30 450 539915
Email: hanna-marie.roehling@charite.de

Abstract

Background: Instrumented assessment of motor symptoms has emerged as a promising extension to the clinical assessment of several movement disorders. The use of mobile and inexpensive technologies such as some markerless motion capture technologies is especially promising for large-scale application but has not transitioned into clinical routine to date. A crucial step on this path is to implement standardized, clinically applicable tools that identify and control for quality concerns.
Objective: The main goal of this study comprises the development of a systematic quality control (QC) procedure for data collected with markerless motion capture technology and its experimental implementation to identify specific quality concerns and thereby rate the usability of recordings.

Methods: We developed a post hoc QC pipeline that was evaluated using a large set of short motor task recordings of healthy controls (2010 recordings from 162 subjects) and people with multiple sclerosis (2682 recordings from 187 subjects). For each of these recordings, 2 raters independently applied the pipeline. They provided overall usability decisions and identified technical and performance-related quality concerns, which yielded respective proportions of their occurrence as a main result.

Results: The approach developed here has proven user-friendly and applicable on a large scale. Raters’ decisions on recording usability were concordant in 71.5%-92.3% of cases, depending on the motor task. Furthermore, 39.6%-85.1% of recordings were concordantly rated as being of satisfactory quality whereas in 5.0%-26.3%, both raters agreed to discard the recording.

Conclusions: We present a QC pipeline that seems feasible and useful for instant quality screening in the clinical setting. Results confirm the need of QC despite using standard test setups, testing protocols, and operator training for the employed system and by extension, for other task-based motor assessment technologies. Results of the QC process can be used to clean existing data sets, optimize quality assurance measures, as well as foster the development of automated QC approaches and therefore improve the overall reliability of kinematic data sets.

(JMIR Hum Factors 2022;9(2):e26825) doi:10.2196/26825

KEYWORDS
instrumented motion analysis; markerless motion capture; visual perceptive computing; quality control; quality reporting; gait analysis

Introduction

With technology rapidly advancing, instrumented motion analysis (IMA) has emerged as an auspicious tool to augment clinical decision-making in persons with motor impairments [1-5]. Applications range from complex gait laboratory equipment to consumer grade health apps, which quantify what a person can do in a standardized setting (motor capacity) or what a person does in everyday life (motor performance) [6].

Regarding motor capacity, marker-based optoelectronic motion analysis systems serve as the gold standard for other technologies [7,8] and are, for instance, successfully used in treatment planning for children with cerebral palsy [9]. However, their high cost and complexity of analysis comprise significant disadvantages for clinical use. Thus, technologies that are portable, affordable, and easy to use are more promising for large-scale application. Respective devices developed for clinical use include pressure-sensitive walkways, inertial sensors (“wearables”), and markerless motion capture systems based on consumer depth cameras [2,10]. In the following, the term IMA will be used for this more versatile subcategory of motion analysis systems.

Despite favorable properties, IMA has not been successfully integrated into wide clinical routine yet [11,12]. Although regulatory requirements for medical products address safety and accuracy within the context of use (eg, for application in specific diseases) [13-15], successful implementation of IMA further depends on acceptance from patients and clinicians. Thus, technical usability, interpretability of outcomes, and quantifiable clinical benefits play a major role in this development. Standardized and efficient quality control (QC) procedures, not only during initial development but also during advancement and application of a system, could facilitate this technological maturation process. We found such QC aspects to be largely understudied and underreported.

QC can be applied at three levels: preventive, ad hoc, and post hoc. Preventive QC is applied before data acquisition. Manufacturers or developing groups generate initial results on data quality and publish them in proof-of-concept studies, including small samples of healthy subjects and target groups for clinical application [7,8,16,17]. Such studies can identify major pitfalls and elaborate on correct usage of these systems. For technology that is already in use with a substantial number of researchers or clinicians, expert consensus can further yield guidelines to improve preventive QC [18]. Ad hoc QC is pertained during measurements. Depending on the system, operators can decide to discard, re-instruct, and rerecord upon observing deviations from standard operating procedures (SOPs) or receiving error messages. Lastly, post hoc QC is employed at the data analysis stage. One option in this context is univariate or multivariate outlier analysis based on the kinematic parameters [19-21]. However, these approaches are highly data-dependent, inept to uncover systematic errors or “false normal” parameter values, and do not provide information regarding underlying causes of data deviation. Additional post hoc QC measures constitute postprocessing tools and successive recalculation of kinematic parameters [22,23] as well as plausibility checks based on raw data [24-26]. To date, such processes have only been performed on comparatively small data sets.

In this study, we used data acquired with the emerging Motognosis Labs system (Motognosis GmbH) that extracts kinematic parameters from depth camera recordings. In recent years, this system was extensively used in a research context at our site and our cooperating sites [24-29] with a standardized protocol for short motor tasks specifically designed to assess motor capacities of people with multiple sclerosis (MS) [7,30]. Regarding preventive QC, previously established SOPs for system operators and patient instructions were used for all data analyzed herein. With respect to ad hoc QC, the software provides visual feedback regarding general subject positioning
in the volume of acquisition and real-time tracking of the whole body as well as individual body parts. Regarding post hoc QC, we found previously employed approaches to be either insufficient, incomplete, or not feasible to reliably examine large amounts of data [19-21,24-26]. Likewise, review of IMA literature did not yield any standards or generalizable concepts. Thus, we propose an approach for systematic post hoc QC, enabling clinical users to prevent, detect, and eliminate data of inferior quality.

For the quality concerns considered here, we distinguish technical and performance issues. Technical issues comprise system-specific malfunctioning of hardware and software as well as artifacts specific to the recording technique, such as signal interference due to subjects’ clothing or the recording environment in the case of depth sensing technology. Performance issues can be considered less technology-specific and can be attributed either to the operator (eg, by providing faulty instructions) or to noncompliance of the recorded subject. If the latter is unrelated to the disease, it should lead to trial exclusion; however, impairment-related inability can be considered a feature of interest.

The main objectives of this study were to (1) build a post hoc QC pipeline that is efficient, user-friendly, and adaptable, enabling clinical users to make standardized and robust decisions concerning usability of individual recordings; (2) perform QC for a large number of recordings acquired at different study sites and thus investigate the types and frequencies of quality issues; and (3) analyze the feasibility of the approach.

**Methods**

**Data Set**

Our study was based on recordings of short, structured motor tasks captured with the Motognosis Labs system. This system relies on a consumer depth camera (Microsoft KinectV2, Microsoft Corporation) and visual perceptive computing. More precisely, the software development kit associated with the camera allows for the markerless tracking of 3D time series from 25 artificial anatomical landmarks for subjects located at 1.5 to 4.5 m from the camera. Custom Motognosis Labs algorithms employ these time series to extract kinematic parameters to quantify various aspects of motor capacity.

Data were pooled from 8 monocentric studies at 3 study sites that used software versions 1.1, 1.4, 2.0, or 2.1 as part of their protocols. These studies will be referred to using the following identifiers: ASD, CIS, Valkinect, VIMS, and WALKIMS-DA (conducted at Charité – Universitätsmedizin Berlin, Berlin, Germany); Ambos and Oprims (conducted at Universitätsklinikum Eppendorf, Hamburg, Germany); and Chiba (conducted at Chiba University, Chiba, Japan). These studies were approved by the respective institutional review boards and all subjects provided written informed consent. The data set comprised recordings from 187 persons with MS and 162 healthy controls. VIMS, Valkinect, and WALKIMS-DA included both groups, whereas the other studies contributed subjects from 1 group only. Descriptive statistics include information on gender, age, anthropometry, and disease severity in case of people with MS, as measured by the Expanded Disability Status Scale [31] (Table 1 and study-specific information in Table S1 in Multimedia Appendix 1).

All subjects performed the Perceptive Assessment in Multiple Sclerosis (PASS-MS) protocol or parts of it between December 2014 and April 2019. PASS-MS consists of 10 structured motor tasks: Postural Control (POCO), Postural Control with Dual Task (POCO-DUAL), Stepping in Place (SIP), Stand Up and Sit Down (SAS), Short Line Walk (SLW), Short Comfortable Speed Walk (SCSW), Short Maximum Speed Walk (SMSW), Pronator Drift Test, Finger-Nose Test, and Finger Tapping. The latter 3 tasks were excluded from this study, as evaluation algorithms were still in an explorative stage at the time, yielding premature claims regarding data quality. A description of the remaining tasks except POCO-DUAL can be found in Otte et al [7,30]. POCO-DUAL equates to POCO with the addition of a cognitive task (Serial 3’s subtraction). System operators had received in-depth training on how to use Motognosis Labs according to written SOPs. System SOPs included specifications of the setup, subject instructions, and rejection guidelines for recordings affected by performance and technical issues. According to the protocol, SAS, SLW, SCSW, and SM SW are recorded thrice consecutively, whereas POCO, POCO-DUAL, and SIP are recorded once. Deviations from SOPs occurred when single tasks or task repetitions were omitted, or operators decided to produce additional recordings (all of which should prompt an operator comment that is stored along with raw data of each recording). Such deviations explain incongruencies in the numbers of recordings per task (Table 1 and study-specific information in Table S2 in Multimedia Appendix 1), as all available recordings were included in this post hoc QC initiative.
Table 1. Demographic information about study subjects with missing data indicated as percentages and number of recordings per Perceptive Assessment in Multiple Sclerosis task subdivided by disease status.

<table>
<thead>
<tr>
<th>Demographic characteristics</th>
<th>All</th>
<th>HC&lt;sup&gt;a&lt;/sup&gt;</th>
<th>PwMS&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N (% female; % —&lt;sup&gt;c&lt;/sup&gt;)</td>
<td>349 (51.6; 0.6)</td>
<td>162 (51.2; 1.2)</td>
<td>187 (51.9; 0)</td>
</tr>
<tr>
<td>Age (years), mean (SD; % —)</td>
<td>42.0 (12.2; 0.6)</td>
<td>38.3 (12.8; 1.2)</td>
<td>45.3 (10.8; 0)</td>
</tr>
<tr>
<td>Height (cm), mean (SD; % —)</td>
<td>173.1 (9.2; 2.6)</td>
<td>172.0 (9.6; 3.7)</td>
<td>174.1 (8.8; 1.6)</td>
</tr>
<tr>
<td>Weight (kg), mean (SD; % —)</td>
<td>72.9 (14.8; 8.0)</td>
<td>70.4 (14.6; 8.0)</td>
<td>75.0 (14.6; 8.0)</td>
</tr>
<tr>
<td>BMI (kg/m&lt;sup&gt;2&lt;/sup&gt;), mean (SD; % —)</td>
<td>24.3 (4.1; 8.0)</td>
<td>23.8 (3.9; 8.0)</td>
<td>24.7 (4.3; 8.0)</td>
</tr>
<tr>
<td>EDSS&lt;sup&gt;d&lt;/sup&gt; median (range; % —)</td>
<td>N/A&lt;sup&gt;e&lt;/sup&gt;</td>
<td>N/A</td>
<td>3.0 (0.0-6.5; 2.7)</td>
</tr>
</tbody>
</table>

# of recordings per PASS-MS<sup>f</sup> task

<table>
<thead>
<tr>
<th>Task</th>
<th>All</th>
<th>HC&lt;sup&gt;a&lt;/sup&gt;</th>
<th>PwMS&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>4692</td>
<td>2010</td>
<td>2682</td>
</tr>
<tr>
<td>POCO&lt;sup&gt;g&lt;/sup&gt;</td>
<td>354</td>
<td>165</td>
<td>189</td>
</tr>
<tr>
<td>POCO-DUAL&lt;sup&gt;h&lt;/sup&gt;</td>
<td>245</td>
<td>88</td>
<td>157</td>
</tr>
<tr>
<td>SCSW&lt;sup&gt;i&lt;/sup&gt;</td>
<td>1043</td>
<td>489</td>
<td>554</td>
</tr>
<tr>
<td>SMSW&lt;sup&gt;j&lt;/sup&gt;</td>
<td>907</td>
<td>361</td>
<td>546</td>
</tr>
<tr>
<td>SLW&lt;sup&gt;k&lt;/sup&gt;</td>
<td>957</td>
<td>428</td>
<td>529</td>
</tr>
<tr>
<td>SIP&lt;sup&gt;l&lt;/sup&gt;</td>
<td>291</td>
<td>131</td>
<td>160</td>
</tr>
<tr>
<td>SAS&lt;sup&gt;m&lt;/sup&gt;</td>
<td>895</td>
<td>348</td>
<td>547</td>
</tr>
</tbody>
</table>

<sup>a</sup>HC: healthy controls.
<sup>b</sup>PwMS: people with multiple sclerosis.
<sup>c</sup>—: not available.
<sup>d</sup>EDSS: Expanded Disability Status Scale.
<sup>e</sup>N/A: not applicable.
<sup>f</sup>PASS-MS: Perceptive Assessment in Multiple Sclerosis.
<sup>g</sup>PFCO: Postural Control.
<sup>h</sup>PFCO-DUAL: Postural Control with Dual Task.
<sup>i</sup>SCSW: Short Comfortable Speed Walk.
<sup>j</sup>SMSW: Short Maximum Speed Walk.
<sup>k</sup>SLW: Short Line Walk.
<sup>l</sup>SIP: Stepping in Place.
<sup>m</sup>SAS: Stand Up and Sit Down.

QC Pipeline Development
The QC pipeline development comprised 2 key components. First, we implemented informative visualizations enabling raters to classify the quality of raw data from PASS-MS recordings and hence implicitly assess the reliability of associated kinematic parameters. Second, we developed an efficient rating strategy for large numbers of recordings.

For the creation of informative visualizations, videos from raw depth streams were generated to enable review of each recorded task. The depth information was further used to produce a condensed representation of each recording in the form of 3 images that are hereafter referred to as motion profiles. They comprise images of depth data averaged over time, over the vertical direction, and over the horizontal direction. As PASS-MS tasks are short and highly standardized, we assumed that major protocol deviations and technical issues would be easily identifiable from motion profiles. To allow for the detection of more subtle quality issues, we also illustrated characteristic signals that are used to calculate kinematic parameters with Motognosis Labs. Visualizations were generated using Python (version 3.7.3) and the matplotlib package (version 3.1.0). A stratified random sample from 15 people with MS and 14 healthy controls was used to test and update visualizations and determine the main rating criteria per task.

We then built a graphical user interface (GUI), which includes a rating window containing visualizations, an overall usability decision checkbox (keep, discard, undecided), and task-specific multiselect checkboxes containing the main rating criteria. Furthermore, on-demand viewers for depth videos and operator comments were integrated. The GUI was programmed in Python (version 3.7.3) using the tkinter package (version 8.6).
prepared detailed rating manuals as well as oral instructions (~45 minutes) to familiarize raters with the GUI. The entire data set (see Table 1) was subjected to ratings, such that each recording was investigated by 2 independent raters. In this step, 8 raters evaluated a total of 4692 recordings from 162 healthy controls and 187 people with MS. Raters comprised medical students, clinician scientists or researchers in other professions, and trained neurologists, all from Charité, Berlin. Among them, 6 raters had operated Motognosis Labs before, whereas 2 were new to the system. Moreover, 2 raters had been actively involved in the development of the QC pipeline, whereas 6 were new to any systematic QC of the data. After in-depth instructions, ratings were conducted individually by the raters at a self-selected speed.

**Statistical Analysis**

Statistical analyses included the extraction of frequencies for overall usability decisions, rater concordance and discordance, and selected rating criteria. The former 2 were illustrated as confusion matrices. Furthermore, the median rating duration per recording was extracted from the GUI log files. Figures were produced with Python (version 3.7.3) using the matplotlib package (version 3.1.0).

**Results**

**QC Pipeline Usage and Feasibility**

After generating visualizations, the implemented GUI can be opened to progressively rate motor task recordings. Intermediate results can be saved in an underlying Excel file, such that raters can flexibly organize their workload. An example of the rating window including respective visualizations, checkboxes, and buttons is shown in Figure 1.

Oral feedback from raters upon completion confirmed that the GUI and the QC pipeline behind it were easy to use and effective. The median rating duration per recording amounted to 6.3 seconds.
Figure 1. Rating window screenshots for an exemplary Stepping in Place recording. Upper left: motion profiles generated by summation of frontally recorded depth data over time, along horizontal and vertical directions and signal curves characteristic of the task (here: knee amplitudes, arm sway, and overall subject positioning over time). Upper right: checkboxes for usability decisions and main criteria including an option for free-text comments. Lower left: on-demand depth video viewer. Lower right: on-demand operator comment viewer.

Rater Concordance and Usability of Recordings

Concerning keep, discard, or undecided decisions, raters concurred on more than 70% of recordings for each task (POCO: 71.5%, POCO-DUAL: 72.7%, SCSW: 92.3%, SMSW: 79.5%, SLW: 74.6%, SIP: 85.6%, and SAS: 90.4%) (Figure 2). Consequently, we observed discordance for up to 28.5% of recordings, which points to task-specific difficulties in using the rating criteria. However, such discordance was mostly due to 1 rater’s undecided decision. Instances of strictly opposing usability, meaning that 1 rater voted keep and the other discard, were uncommon (between 0.8% and 4.9%), except for SMSW (10.5%).
Figure 2. Synopsis of usability decisions by 2 raters per recording per Perceptive Assessment in Multiple Sclerosis task. Rater agreement on usability decisions keep, discard, and undecided are framed. POCO: Postural Control; POCO-DUAL: Postural Control with Dual Task; SAS: Stand Up and Sit Down; SCSW: Short Comfortable Speed Walk; SIP: Stepping in Place; SLW: short line walk; SMSW: Short Maximum Speed Walk.

A task-wise visualization of rater decisions regarding usability of recordings is depicted in Figure 2. Unobjectionable usability, defined as a unanimous keep decision, was obtained for 85.1% of SCSW, more than 70% of SMSW and SIP (73.3% and 70.8%, respectively), more than 60% for SAS and SLW (62.9% and 60.5%, respectively) and less than or close to half for POCO and POCO-DUAL recordings (50.3% and 39.6%, respectively). The highest rates for unanimous discard decisions were observed for SAS (26.3%), followed by POCO-DUAL (25.3%), and POCO and SIP (13.0% and 13.1%, respectively). The respective rates were low for gait tasks including SLW, SCSW, and SMSW (9.4%, 6.5%, and 5.0%, respectively). Rater concordance as well as proportions of unanimous keep and discard decisions subdivided for all studies can be found in Table S3 in Multimedia Appendix 1.

Main Quality Concerns

The main rating criteria compiled during QC pipeline development are listed below, with the respective tasks indicated in parentheses.

- Disturbances, technical issue: Signal disturbances including noisy background, floor, and technical issues with tracking clothing (all tasks)
- Duration, technical issue: Recording duration substantially deviating from 40 seconds, namely a deviation of more than 1 second (POCO, POCO-DUAL, and SIP)
- Step Detection, technical issue: Incorrect Step Detection (SCSW, SMSW, SIP, and SLW)
- Up/Down Phase, technical issue: Incomplete or incorrectly detected standing-up or sitting-down phase (SAS)
- Arms, performance issue: Arms not hanging loosely down at the beginning of the recording (SAS)
- Backward, performance issue: Subject walking backward by more than 50 cm or exhibiting a deliberate backward correction (SIP)
- Feet, performance issue: Deviation from closed feet position, namely if the feet are in an open or a V-shaped position (POCO and POCO-DUAL)
- Forward, performance issue: Subject moving forward by more than 50 cm (SIP)
- Movements, performance issue: Task-unassociated movements such as scratching or gesturing (POCO, POCO-DUAL, SLW, SIP, and SAS)
- Sidestep, performance issue: 1 or multiple sidesteps (POCO, POCO-DUAL, and SLW)
- Support, performance issue: Subject needing support from a walking stick, walls, rollator, or the like (all tasks)
- Other, technical or performance issue: Other/unlisted criterion (all tasks)

Respective selection frequencies (multiple selections were possible) are illustrated in Figure 3. Possible disease-associated differences in data quality can be estimated from the 3 studies featuring healthy controls and people with MS, namely VIMS, Valkinect, and WALKIMS-DA.
The most prevalent quality concerns comprised Feet, Disturbances, and Other for POCO and additionally Movements for POCO-DUAL. An example of a POCO recording that was discarded due to incorrect Feet positioning as well as unassociated Movements, namely the most frequent performance-associated quality concerns, can be found in Figure 4. For POCO-DUAL, supposedly task-unassociated movements were tagged with Movements and Other by the raters. However, these hand and arm movements often seemed to result from cognitive efforts made during mental arithmetic. In this case, no clear distinction between task-associated and task-unassociated movements can be made. Regarding technical quality concerns, raters’ comments suggested that recordings tagged with Disturbances or Other most often exhibited noisy or corrupt leg, feet, or floor signals.
Prevalent quality concerns for gait tasks were Disturbances and Step Detection in SLW and— less frequently—SCSW and SMSW. A cross-dependency between the 2 criteria was often observed when unsuitable clothing led to noisy signals (noted as Disturbances by the raters), which in turn leads to issues concerning Step Detection. An example of this issue for an SCSW recording is depicted in Figure 5. Other Disturbances related to floor reflections were not associated with Step Detection issues as often.

Excessive forward locomotion (Forward) was the most frequent quality concern for SIP recordings. However, from our experience, the chosen threshold of 50 cm forward motion is rather conservative and distances up to 80-100 cm might be tolerable.

The most prominent problem for SAS was incorrect arm positioning (Arms) at the beginning of a recording. Such incorrect arm positioning was not easily discernible from the motion profile alone and raters usually consulted the provided depth videos to confirm this specific quality concern. Furthermore, a mistake in signal plot generation for
SAS—ffecting 3.8% of SAS plots—led to an overestimation of recordings affected by the Up/Down Phase criterion. Figure 3 provides raw ratings, and the represented numbers hence reflect this overestimation.

Disparities between people with MS and healthy controls for performance-related quality aspects were apparent for the generally less often observed Support (all tasks) and Sidestep (POCO, POCO-DUAL, and SLW) issues. This can be interpreted as a disease-related difficulty or the inability to follow task instructions. Results regarding incorrect Feet positioning during POCO and POCO-DUAL did not allow for the interpretation of this criterion as a mainly disease-related one. This criterion as well as Forward and Backward motion during SIP and the incorrect starting position of the Arms during SAS were present in both groups, though slightly more frequent in people with MS. Frequencies of the observed quality criteria further subdivided for all studies can be found in Table S4 in Multimedia Appendix 1.

**Discussion**

This study presents a post hoc QC pipeline for clinical users of an IMA system. Its core consists of an interface, which enables an intuitive usability decision for individual recordings based on an extendable set of quality criteria. The pipeline proved highly feasible for users—including raters less acquainted with the IMA system itself—and yielded acceptable rater concordance. Its application in a large set of recordings from healthy controls and people with MS demonstrated the utility and necessity of post hoc QC to ensure reliable data and avoid misinterpretation of IMA results. It further identified points for improvement in preventive and ad hoc QC. To our knowledge, this is the first study to systematically investigate QC aspects and propose a clinically applicable QC pipeline for visual perceptive computing.

In the following, we will discuss 2 main aspects of our results. First, the rater concordance, which indicates the feasibility and limitations of our QC approach, and second, the usability decisions themselves, which indicate the quality and limitations of our data.

Rater concordance between 71.5% to 92.3% was generally acceptable. Only for SMSW, strictly opposed keep/discard decisions occurred to a relevant extent (10.5%). This was mostly caused by 1 rater’s discard decisions because no full gait cycle was captured. Due to the limited recording range of the depth camera, this is a frequent observation for SMSW and cannot be directly attributed to technical or performance issues. Generally, discordance may reflect ambiguity regarding rating criteria, difficulties in the evaluation of individual cases, or rater oversight. Probably only 1 rater, most likely the operator of the system, will apply post hoc QC in future clinical applications. Thus, possible reasons for rater discordance should be carefully addressed in further development of the QC pipeline, for instance, by specifying the rating criteria, as well as conducting more targeted rater trainings. However, as with other clinical judgments, QC decisions will remain informed, but ultimately intuitive decisions.

Usability decisions were interpreted as follows. Recordings receiving a unanimous keep or discard decision from the corresponding 2 raters were regarded as having assessable and satisfactory or unsatisfactory quality, respectively. Remaining recordings with discordant or undecided usability decisions were classified as needing further investigation, thus being less assessable and with potentially objectionable quality. The proportion of unanimous keep decisions varied substantially between tasks (39.6%–85.1%). In this respect, the SCSW task had the most favorable results with the highest rater concordance (92.3%) and the highest proportion of keep decisions among all tasks. At the other end of the spectrum were POCO and POCO-DUAL with rather moderate rater concordance (71.5% and 72.7%, respectively) and comparatively less unanimous keep decisions (50.3% and 39.6%, respectively). This partial ambiguity supports our inclusion of undecided as an option to avoid forced decisions as well as free text comments to enable marking of unexpected quality concerns.

Regarding technical quality issues, the short walk tasks SCSW, SMSW, and SLW suffered the most from unfavorable properties of clothing that hampered infrared light reflection [32]. POCO and POCO-DUAL often exhibited noisy and cutoff feet signals, attributable to a limited differentiation of feet and ground leading to unstable landmark estimations, as reported earlier [7]. Countermeasures include general recommendations toward subjects’ clothing and flooring at the measurement site.

We expected performance-related quality concerns to be associated with physical limitations and thus the disease status to some extent. This seemed to apply to rating criteria Sidestep and Support. However, the more commonly observed performance-related issues (eg, Feet and Movements for POCO and POCO-DUAL, Arms for SAS, and Forward for SIP) occurred in healthy subjects as well. This implies that mistakes in task instruction or ad hoc QC occurred to a relevant degree, despite detailed SOPs and operator training. Even higher proportions of performance-related issues may be expected with wider clinical use or in unsupervised telemedical applications. Thus, further IMA development should aim to implement technical measures for automated real-time detection of performance issues and respective response plans (eg, reinstruction and repetition). Performance-related quality concerns may specifically apply to the assessment of motor capacity in a lab setting or in task-based assessments as opposed to the recently proposed IMA systems for continuous assessment of motor performance [4,5,15].

In the literature, we found generally sparse reporting of QC aspects for IMA. This includes reporting of unobjectionable data quality, which we assume to be unlikely. As an indicator of technical IMA system performance, some authors reported exclusion of IMA recordings due to seemingly blatant technical failures, with rates ranging from a few corrupted examples to recordings of 48.8% of the participants [21,22,33,34]. Unfortunately, respective proportions could not be provided for our data set, as we did not track recordings discarded ad hoc. Regarding data exclusion in postprocessing, outlier detection was the most frequent approach. For univariate outlier detection on normative gait and balance parameters in children, exclusion rates of 2.5% and 6% were reported [20,35]. A multivariate
outlier detection approach on kinematic gait data with successive expert evaluation identified erroneous Step Detection in 3.4% of the subjects [21], whereas a custom post hoc QC procedure applied on SMSW data obtained using Motognosis Labs led to exclusion of 6.7% of the recordings [24]. We consider the QC approach presented here to be rather conservative when compared to outlier detection. It is highly possible that significant quality concerns identified at the raw data level would not be detected by outlier analysis at the kinematic parameter level. For example, failure to stand with closed feet during POCO most likely results in reduced postural sway, which would be mistaken for higher postural stability in the respective subject at the kinematic outcome level.

Lastly, reporting of manual postprocessing, for example, using the GAITRite footfall labeling tool, is often limited to whether it was employed at all [22,36], and respective proportions are only seldom addressed [37].

Beyond IMA, the need for QC has been recognized for other technical procedures. In the context of MS research, magnetic resonance imaging and optical coherence tomography serve as examples for which recommendations have been made regarding standardized protocols, QC, and harmonious reporting thereof [38-42]. Therefore, we propose standardized reporting of IMA results to include information regarding the following: (1) number of recording failures during data acquisition; (2) type and amount of applied postprocessing, both technical and manual; (3) fraction of recordings undergoing QC; (4) fraction of recordings ultimately excluded from analysis (mention of respective causes would be highly valuable for future users).

Limitations of this study may include the decision to have each recording viewed by 2 out of 8 available raters; this limits formal interrater reliability analyses and does not assess individual rater bias. However, we did not aim to establish interrater reliability but focused on obtaining generalizable estimates of rater concordance and determining the feasibility of the approach with a reasonably diverse set of raters. Further, other possible factors influencing usability of the recordings were not specifically analyzed. These include effect of the study site, population, system operators, as well as subjects’ age, height, and weight. However, we consider QC results generalizable to and representative of routine applications because of the large size and heterogeneity of our sample. Differences in hardware were not tracked in this study (Kinect 2 sensors and laptops). Likewise, differences in software versions were disregarded because they were considered not substantial. However, recommendations regarding hardware and software may prospectively play a role in preventive QC in large-scale applications.

Regarding transferability, the visualizations employed here were specific to Motognosis Labs. However, appropriate visualizations have been implemented for other IMA systems as well. Examples include footprint depictions from pressure-sensitive walkways or acceleration illustrations from inertial sensors. Thus, we expect the general QC approach presented in this study to be transferable to other IMA systems. As for the observed quality concerns, technical issues are mostly or partially transferable to other depth camera– or visual sensor–based systems, respectively. The performance issues observed here are even more generalizable and thus highly informative for all researchers and clinicians using lab- or task-based IMA. The results of this study clearly support the need for QC of IMA data to ensure objectivity and enhance acceptance by clinical users and regulators alike. As a first step, this approach can advance consensus on the QC standards of different IMA systems and ultimately improve data quality.

Acknowledgments

We thank NeuroCure Clinical Research Center (NCRC), funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy – EXC-2049 – 390688087 and Charité-BIH Clinical Study Center. Parts of this work were funded by a restricted research grant from Roche. DK is a participant in the BIH-Charité Clinician Scientist Program funded by the Charité –Universitätsmedizin Berlin and the Berlin Institute of Health. SCR was supported by the Clinician-Scientist Fellowship from the Stifterverband für die Deutsche Wissenschaft and the Hertie Network of Excellence in Clinical Neuroscience of the Gemeinnützige Hertie-Stiftung.

Conflicts of Interest

aub is a shareholder and hmr is a paid part-time employee at motognosis gmbh. the other authors declare no conflict of interest.

Multimedia Appendix 1

Further information on data set, usability decisions, and rating criteria statistics.

[DOCX File, 110 KB - humanfactors_v9i2e26825_app1.docx]

References


Abbreviations

GUI: graphical user interface  
IMA: instrumented motion analysis  
MS: multiple sclerosis  
PASS-MS: Perceptive Assessment in Multiple Sclerosis (ie, name of short motor assessment battery recorded with Motognosis Labs)  
POCO: Postural Control  
POCO-DUAL: Postural Control with Dual Task  
QC: quality control  
SAS: Stand Up and Sit Down  
SCSW: Short Comfortable Speed Walk  
SIP: Stepping in Place  
SLW: Short Line Walk  
SMSW: Short Maximum Speed Walk  
SOP: standard operating procedure
Evaluation of the First Year(s) of Physicians Collaboration on an Interdisciplinary Electronic Consultation Platform in the Netherlands: Mixed Methods Observational Study

Sanavro et al

Abstract

Background: Complexity of health problems and aging of the population create an ongoing burden on the health care system with the general practitioner (GP) being the gatekeeper in primary care. In GPs daily practice, collaboration with specialists and exchange of knowledge from the secondary care play a crucial role in this system. Communication between primary and secondary care has shortcomings for health care workers that want to practice sustainable patient-centered health care. Therefore, a new digital interactive platform was developed: Prisma.

Objective: This study aims to describe the development of a digital consultation platform (Prisma) to connect GPs with hospital specialists via the Siilo application and to evaluate the first year of use, including consultations, topic diversity, and number of participating physicians.

Methods: We conducted a mixed methods observational study, analyzing qualitative and quantitative data for cases posted on the platform between June 2018 and May 2020. Any GP can post questions to an interdisciplinary group of secondary care specialists, with the platform designed to facilitate discussion and knowledge exchange for all users.

Results: In total, 3674 cases were posted by 424 GPs across 16 specialisms. Most questions and answers concerned diagnosis, nonmedical treatment, and medication. Mean response time was 76 minutes (range 44-252). An average of 3 users engaged with each case (up to 7 specialists). Almost half of the internal medicine cases received responses from at least two specialisms in secondary care, contrasting with about one-fifth for dermatology. Of note, the growth in consultations was steepest for dermatology.

Conclusions: Digital consultations offer the possibility for GPs to receive quick responses when seeking advice. The interdisciplinary approach of Prisma creates opportunities for digital patient-centered networking.

(KEYWORDS primary care; digital consultation; interdisciplinary; specialist care

doi:10.2196/33630)
Introduction

In the Dutch health care system, general practitioners (GPs) have a coordinating role as generalists, functioning as gatekeepers to secondary care. This model requires that patients initially consult a GP who provides expert generalist medical care for their health care problem and considers the need for referral to more specialist care.

Unfortunately, pressures on the health care system have increased due to the growth in both the chronicity and the complexity of health problems [1,2]. Although GPs care for over 95% of medical problems that present during consultations, referral to secondary care has also increased, resulting in greater health care costs and growing waiting lists [3,4]. These issues can be addressed by providing GPs with closer support from secondary care, assuming there are effective routes for knowledge exchange [5-9]. However, the most commonly used tools for communication between primary and secondary care have important shortcomings. For example, GPs and hospital specialists are often mutually unavailable at the same time, meaning that telephone conversations can be interruptive. Whereas e-consultations may solve the problem of asynchronous availability, they are limited by being monodisciplinary, one-on-one, and mostly noninteractive [6,10-13]. Digital response times may also vary by specialism. By contrast, team-based case collaboration on a patient-centered network of health care professionals could facilitate communication and knowledge transfer [14-16]. The secure Siilo app offers a useful platform to host such a service [17,18].

In this study, we describe the development of the Dutch Prisma platform within the secure Siilo app and evaluate the usage and consultations in the first 2 years since its introduction, including the diversity of topics and number of physicians involved.

Methods

Study Design

We performed a retrospective mixed methods study using quantitative information from the Prisma platform and a qualitative evaluation of consecutive cases posted on the platform from its inception in July 2018 to May 2020.

The Prisma Platform

The Prisma platform initially facilitated digital interprofessional consultation for patients with orthopedic problems, but more recently, it has expanded to include other specialties. GPs with full access to the closed digital environment of the platform generate cases by providing anonymized patient information with a question. All GPs and specialist users are connected in so-called tiles by specialty (eg, orthopedics, internal medicine, palliative care) to facilitate engagement by consultants with complementary expertise (eg, rheumatologists, orthopedic surgeons, sports medicine physicians, and radiologists participate via the orthopedics tile). All users can engage with each tile and upload attachments or links to relevant information, such as laboratory results, pictures, or guidelines. The main language used on the platform is Dutch.

Two GP groups are active on this platform: 1 with full access (able to generate cases and respond to others) and 1 with a read-only account. Specialists participated voluntarily; separate from their hospital work and without reimbursement for their activity on the platform. Because they were not reimbursed, the number of GPs was limited during the development phase to avoid overloading the specialists. All users, both GPs and specialists, were located in various regions of the Netherlands. Specialists preferably respond within 24-48 hours by answering questions, seeking more information, or engaging in discussion. All GPs with access to the platform can read and respond to posted cases. In this way, the platform allows for a dynamic exchange of information and learning to support the GP in daily practice. Throughout the process, the GP remains responsible for the care provided to the patient and will decide, in consultation with the patient, how to proceed with further treatment.

Data Collection

A data analyst at Siilo provided pseudonymized details for all consecutive cases, replacing usernames with a job title and a number (eg, GP-1, GP-2, neurologist-1). Each post was summarized as a user code, timestamp, and verbatim transcript, and these were grouped by case for each tile. Data were analyzed qualitatively and quantitatively. As we performed a retrospective descriptive study, we did not predefine our sample size.

Qualitative Analysis

Text files were imported into the Atlas.ti program [19] for qualitative assessment by a research group comprising 20 senior medical students (coders) supervised by an internist (SS), a medical sociologist (DJ), a GP epidemiologist (MB), and a senior researcher (HW). The Prisma affiliate (PK) was not involved in this phase.

We used a predefined coding tree to structure the qualitative assessment (Multimedia Appendix 1). Before applying this to all cases, a random sample of 10 cases was initially coded by all coders. The results of this preliminary coding were then checked in pairs and discussed in 5 subgroups with 2 supervisors. Coders were actively invited to discuss the applicability of codes and to add new codes if needed. After this, coders were grouped by tile and at least 50 cases per tile were coded in duplicate with mutual blinding. This was followed by group discussion in consensus meetings per subgroup, after which the remaining cases were coded.

The coding tree comprised the following: basic patient characteristics, such as age, gender, and comorbidity; the topic of the question; and both the type of question and the type of answer (eg, diagnostic, therapeutic, or referral for both). Codes for symptoms and diseases followed the International Classification in Primary Care (ICPC) [20], with multiple codes permitted.

Quantitative Analysis

All codes were imported into IBM SPSS (IBM Corp.) for quantitative analysis. We merged the 16 tiles into 5 categories based on similarities and group sizes: “internal medicine” included internal medicine, infectious disease, palliative care,
and medically unexplained physical symptoms; “observation” included gastroenterology, neurology, pulmonology, rheumatology, and cardiology; “surgical” included orthopedics, urology, traumatology, and ear, nose, and throat disease; “female/child” included gynecology and pediatrics; and “dermatology” as a single category. The tile for psychiatry was analyzed and published separately and is therefore excluded from this analysis [21].

An overview of activity on the platform is displayed by plotting the number of GPs (active users and read-only accounts) and the number of cases against time. We estimated the number of users, number of specialisms, number of specialists, and the response time for each case based on user codes and timestamps, and we analyzed the code frequencies for age, gender, case topic (based on the ICPC code), question type, and answer type for each category. Descriptive data were presented as percentages of all cases or as means and SDs. Finally, we used a Sankey diagram to show the linkage between questions and answers.

Results

Descriptive Data

The data set started with 25,954 messages for 4013 cases; of these, 1872 messages were excluded for 339 cases. First, we excluded 292 cases because of data extraction errors (n=34), small size, and difficulty to categorize within groups (geriatrics, n=5; ophthalmology, n=40) and because they were already analyzed in a separate study (psychiatry, n=213) [21]. Next, we divided the data within the research team and analyzed the 3721 cases. We excluded another 47 cases because of wrong tile placement (n=19), double case placement (n=10), technical errors (n=8), not coded (n=7), withdrawal by GP (n=2), missing (n=1) (Multimedia Appendix 2). The 3674 included cases were posted by 424 different GPs (median 9 cases per GP), for whom 97 (22.9%) first posts were in response to another case and 327 (77.1%) posts were for new cases.

Growth of the Prisma platform over time is shown as the number of GPs (active users and read-only accounts; Figure 1), the total number of cases, and the number of cases per tile (Figure 2). The number of cases per category was 677 for internal, 674 for observation, 860 for surgical, 875 for female/child, and 588 for dermatology. Figures 3 and 4 show the number of specialists and specialisms involved per tile category, respectively. For all categories, except dermatology (196/588, 33.3%), most cases included more than 2 users per case. For the internal, observation, and surgical categories, 3 or more specialisms were involved per case in 46.6% (317/680), 32.3% (217/672), and 40.7% (350/860), respectively. In the internal and observation categories, 4 or more health care professionals were engaged per case in 57.2% (389/680) and 54.0% (363/672), respectively.

Figure 1. Platform use; number of active and read-only GPs on the platform. GP: general practitioner.
Figure 2. Overall cases of network activity and network activity by tile category. ENT: ear, nose, throat; GYN: gynaecology; MUPS: medically unexplained physical symptoms; PAL: palliative care; UROL: urology.

Figure 3. Number of users involved per case. Data are illustrated in 5 tile categories.

Figure 4. Number of specialisms involved per case. Data are illustrated in 5 tile categories.
Case characteristics are presented in Multimedia Appendix 3. No answer was given for 35 cases, with the median time to first response being 76 minutes (IQR 17-320) for the other cases. The shortest response time was seen in the surgery category (median 44 minutes) and the longest was in the dermatology tile (median 252 minutes). Overall, 3508/3674 (95.48%) cases contained specific patient information or patient-specific questions, with the remaining 166 (4.52%) cases including questions that were not specific to the patient. Slightly more than half of all queries concerned females (1948/3674, 53.02%), except for those in the surgical tile where there was a slight male majority (437/860, 50.8%). GPs did not report gender in 8.92% (313/3508) of the patient-specific cases. They also posted a question about more than 1 patient in 4 cases (eg, family members or several patients with the same complaint). Patient age ranged from newborn to 101 years (mean 39.9 years) and the mean age differed by tile category. The GP did not report age for 701 cases.

Topics discussed covered the full range of ICPC codes (Multimedia Appendix 4). The 3 main topics by ICPC code were in the skin, musculoskeletal, and general symptom domains.

Type of Questions and Answers
Among the 3674 cases, we identified 6691 different questions (mean 1.8 per case) and 10,922 answers (mean 3.03 per case).

Multimedia Appendix 5 shows the type of question and answers posted.

Questions concerned (differential) diagnosis in 50.90% (1870/3674), appropriate nondrug treatment in 33.15% (1218/3674), and drug treatment in 27.60% (1014/3674). It was notable that the focus of questions differed between tile categories. Most concerned diagnosis in the internal (358/677, 52.9%), observatory (361/674, 53.6%), and dermatology (424/588, 72.1%) categories; most concerned treatment in the surgical category (431/860, 50.1%); and most concerned medication in the female/child category (378/875, 43.2%).

The Sankey diagram in Figure 5 illustrates the dynamics between the type of question and the type of answer. We have illustrated only the 9 most common combinations (used more than 100 times), including any other answer type or combination in the “other” group. Consistent with the type of question asked by GPs, most answers concerned (differential) diagnosis, which was often combined with responses about referral, further diagnostics, or a combination of these 3 responses. However, the type of question posed by GPs did not always lead to answers within the same topic, such as questions about referral often leading to advice about how to proceed (eg, perform further diagnostics and refer, GP-based follow-up, or start therapy and refer). In this way, one can see that simple referral questions can lead to varied advice possibilities (Multimedia Appendices 6-8).
Discussion

Principal Findings

This mixed methods study has shown the growth and evolution of a digital interdisciplinary consultation platform over almost 2 years. Posted questions not only covered a broad spectrum of the population by age and sex but also covered a wide variety of specialist topics. Of note, there was a steep increase in the number of cases for dermatology, which could be explained by existing familiarity with tele-dermatology in Dutch primary care [10] or potentially highlight a practice weakness among GPs.

In most cases, 2 or more users engaged with the GP who initiated the question. An exception to this was the dermatology tile, in which it was typical for only 1 other user to respond. The number of involved specialisms also differed between tiles, being largest for internal medicine. This illustrates a novelty of this approach compared with other consultation formats where a GP only has contact with 1 medical specialist. This approach is in line with the future vision to build primary and secondary care networks around the patient [16,22,23].

The short response times suggest that the Prisma platform facilitates rapid and efficient consultation. This contrasts with telephone consultations, which are often hampered by mutual unavailability. Our data indicate that answers are given to most questions by the end of a GP’s working day so that patient care is not delayed for more than a few hours.

Although it is difficult to compare our study with previous studies because of the difference in design of the platform that was analyzed, the time response outcomes are superior to those in previous studies [4,6,11,24]. It should be considered that they may reflect a precursor effect of enthusiasm among engaged specialists.

The differences in question type between tile categories may indicate differences in work content. Internal medicine, observation, and dermatology focused on diagnosis; surgery focused on treatment; and female/child focused on medication. An alternative hypothesis could be that different specialisms have specific needs of GPs in the treatment process.

The Sankey graph in Figure 5 and Multimedia Appendices 6-8 illustrates the dynamics between questions asked by GPs and answers given by specialists. The large number of questions related to diagnosis had multiple combinations with other questions, reflecting the complexity of evaluation (eg, when the diagnosis is unclear, the next step is also uncertain). Overall, (differential) diagnosis was the most frequently used theme, but this does not appear as a separate group in the graph because it was mostly used in combination with other themes. In comparison to this, questions on medication had most single questions and a clear dynamic to single answers.

The dynamics on referral questions are also interesting, with only a minority of questions receiving a single answer about referral. For example, we found combinations of advice for additional diagnostics in primary care or advice to refer with explanations about diagnosis. We hypothesize that medical specialists used this platform not only to ensure adequate referral but also to share knowledge. There was also a difference between referral questions and answers: not all questions about referral led to answers about referral, and vice versa (ie, referral advice was sometimes given without a specific request).

We found similarities and differences when comparing our findings with the limited amount of preceding research on electronic consultations [15]. In this earlier research, most questions for hematology and rheumatology concerned diagnosis, while questions in the infectious disease and dermatology categories typically concerned therapy. Another research focusing specifically on internal medicine in a hospital in Netherlands involved one-on-one electronic consultations, and revealed “diagnostic tools” to be the most common answer [6].

Limitations

First, the large sample size and categorization means that a more detailed analysis by specialty is missing in this study. Second, because structure was lacking in the questions posted by GPs, complete data on patient characteristics cannot be guaranteed; however, this did not impair the content analysis. Third, text coding was done by 20 different coders, which might have resulted in interobserver variations in interpretation, despite our efforts to minimize this as much as possible through teamwork. Finally, the data in this analysis were observational in nature, preventing us from making firm conclusions on either observed correlations or patient outcomes.

Future Research

This evaluation focused on the activities of health care professionals, but to date, the patient perspective has not been analyzed. Although the platform performs well in supporting the needs of the GP for further assessment, treatment, and when needed, more appropriate referral to specialists, we do not know how these relate to needs, experiences, and outcomes in patient cohorts. To generate and implement a novel health care collaboration on a large scale, time and cost-efficiency calculations will also be indispensable [25]. In our study, the response time was more rapid than previously reported for e-consultations [6,24], which have already been shown to reduce not only waiting times for GPs and patients but also costs for patients and waiting lists for hospitals [26]. We are currently conducting a stepped-wedge randomized controlled trial to evaluate the impact of the Prisma platform on patient outcomes and referrals to specialists.

Concerning the content of questions posted on the Prisma platform, an in-depth analysis could still be interesting and useful. Gaps in support for GPs could be uncovered by exploring diagnostic uncertainties (between noncomplex symptoms that meet ICPC diagnostic criteria and practice guidelines), common reasons for referral, and the impact of regional agreements [27]. It is possible that these gaps could be filled by creating a database of information collected on the platform. This could facilitate GPs to ask questions and search for possible answers based on prior responses.
**Conclusion**
This observational research shows that a new digital platform facilitated rapid and interactive communication between GPs and specialists for nonurgent questions. This platform is clearly distinguished from one-to-one consultations by facilitating the involvement of multiple physicians. The platform supports the transfer of knowledge from medical specialists to GPs while allowing different viewpoints from relevant experts.

**Acknowledgments**
We are grateful to the Department of General Practice, University Medical Centre Groningen, and to Prisma for providing technical support. Prisma Platform Study Group: Ineke Knijp, Ruben BR de Boer, Lisa Havinga, Iris R Vroom, Esmeee L van der Geest, Maaike A Hulshof, Daphne BM Visser, Frank J Dorgelo, Ivar Maatje, Hugo Quaedvlieg, Lianne Sijbring, Anne van der Meer, Esmeee Fath el Bab, Lucinda E Haaze, Willeke Schelhaas, Susan JM Bergamin, Kimberly Boerma, Anne J Lammers, Dajana Erceg, Tamara Schouten, and Steven N Koning. We thank Dr Robert Sykes for providing technical editing services for the final drafts of this manuscript.

**Conflicts of Interest**
PK is the founder of Prisma and provided the text messages of the platform for this research, but was not involved in the data analysis process of this work.

**Multimedia Appendix 1**
Code tree (in Dutch).
[PDF File (Adobe PDF File), 165 KB - humanfactors_v9i2e33630_app1.pdf ]

**Multimedia Appendix 2**
Flowchart inclusion cases.
[PNG File, 24 KB - humanfactors_v9i2e33630_app2.png ]

**Multimedia Appendix 3**
Characteristics of cases posted in different tiles.
[DOC File, 37 KB - humanfactors_v9i2e33630_app3.doc ]

**Multimedia Appendix 4**
International Classification in Primary Care codes used in cases.
[DOC File, 47 KB - humanfactors_v9i2e33630_app4.doc ]

**Multimedia Appendix 5**
Type of questions and type of answers.
[DOC File, 53 KB - humanfactors_v9i2e33630_app5.doc ]

**Multimedia Appendix 6**
[PNG File, 1867 KB - humanfactors_v9i2e33630_app6.png ]

**Multimedia Appendix 7**
Sankey diagram for drug treatment questions.
[PNG File, 2742 KB - humanfactors_v9i2e33630_app7.png ]

**Multimedia Appendix 8**
Sankey diagram for referral questions.
[PNG File, 390 KB - humanfactors_v9i2e33630_app8.png ]

**References**

https://humanfactors.jmir.org/2022/2/e33630

JMIR Hum Factors 2022 | vol. 9 | iss. 2 | e33630 | p.74

(page number not for citation purposes)


Abbreviations

GP: general practitioner
ICPC: International Classification in Primary Care

©Sanne M Sanavro, Henk van der Worp, Danielle Jansen, Paul Koning, Marco H Blanker, Marco Prisma Platform Study Group. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 01.04.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Implementation of a Web-Based Tool for Shared Decision-making in Lung Cancer Screening: Mixed Methods Quality Improvement Evaluation

Julie Lowery1*, PhD; Angela Fagerlin2,3, PhD; Angela R Larkin1*, BA; Renda S Wiener4,5, MPH, MD; Sarah E Skurla1*, MPH; Tanner J Caverly1,6,7*, MPH, MD

1Center for Clinical Management Research, Ann Arbor VA Healthcare System, Ann Arbor, MI, United States
2Department of Population Health Sciences, University of Utah School of Medicine, Salt Lake City, UT, United States
3Informatics Decision-Enhancement and Analytics Sciences Center for Innovation, VA Salt Lake City Healthcare System, Salt Lake City, MI, United States
4Center for Healthcare Organization & Implementation Research, VA Boston Healthcare System, Boston, MA, United States
5The Pulmonary Center, Boston University School of Medicine, Boston, MA, United States
6Department of Learning Health Sciences, University of Michigan School of Medicine, Ann Arbor, MI, United States
7Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States

*these authors contributed equally

Corresponding Author:
Tanner J Caverly, MPH, MD
Center for Clinical Management Research
Ann Arbor VA Healthcare System
2215 Fuller Road
Ann Arbor, MI, 48105
United States
Phone: 1 303 587 1038
Email: tcaverly@med.umich.edu

Abstract

Background: Lung cancer risk and life expectancy vary substantially across patients eligible for low-dose computed tomography lung cancer screening (LCS), which has important consequences for optimizing LCS decisions for different patients. To account for this heterogeneity during decision-making, web-based decision support tools are needed to enable quick calculations and streamline the process of obtaining individualized information that more accurately informs patient-clinician LCS discussions. We created DecisionPrecision, a clinician-facing web-based decision support tool, to help tailor the LCS discussion to a patient’s individualized lung cancer risk and estimated net benefit.

Objective: The objective of our study is to test two strategies for implementing DecisionPrecision in primary care at eight Veterans Affairs medical centers: a quality improvement (QI) training approach and academic detailing (AD).

Methods: Phase 1 comprised a multisite, cluster randomized trial comparing the effectiveness of standard implementation (adding a link to DecisionPrecision in the electronic health record vs standard implementation plus the Learn, Engage, Act, and Process [LEAP] QI training program). The primary outcome measure was the use of DecisionPrecision at each site before versus after LEAP QI training. The second phase of the study examined the potential effectiveness of AD as an implementation strategy for DecisionPrecision at all 8 medical centers. Outcomes were assessed by comparing absolute tool use before and after AD visits and conducting semistructured interviews with a subset of primary care physicians (PCPs) following the AD visits.

Results: Phase 1 findings showed that sites that participated in the LEAP QI training program used DecisionPrecision significantly more often than the standard implementation sites (tool used 190.3, SD 174.8 times on average over 6 months at LEAP sites vs 3.5 SD 3.7 at standard sites; P<.001). However, this finding was confounded by the lack of screening coordinators at standard implementation sites. In phase 2, there was no difference in the 6-month tool use between before and after AD (95% CI −5.06 to 6.40; P=.82). Follow-up interviews with PCPs indicated that the AD strategy increased provider awareness and appreciation for the benefits of the tool. However, other priorities and limited time prevented PCPs from using them during routine clinical visits.
Conclusions: The phase 1 findings did not provide conclusive evidence of the benefit of a QI training approach for implementing a decision support tool for LCS among PCPs. In addition, phase 2 findings showed that our light-touch, single-visit AD strategy did not increase tool use. To enable tool use by PCPs, prediction-based tools must be fully automated and integrated into electronic health records, thereby helping providers personalize LCS discussions among their many competing demands. PCPs also need more time to engage in shared decision-making discussions with their patients.

Trial Registration: ClinicalTrials.gov NCT02765412; https://clinicaltrials.gov/ct2/show/NCT02765412

KEYWORDS
shared decision-making; lung cancer; screening; clinical decision support; academic detailing; quality improvement; implementation

Introduction

Background

National lung cancer screening (LCS) guidelines have consistently recommended shared decision-making (SDM) to inform patients about the pros and cons of low-dose computed tomography (LDCT) screening and individualized LCS decisions [1,2]. The Centers for Medicare and Medicaid Services require documentation of SDM before initiating LDCT screening for its covered population, a policy that is unique among screening decisions in primary care [3]. Thus, understanding how to best implement and optimize SDM for LCS has been an urgent task for all health systems and clinicians offering LCS to their eligible patient population.

A key approach to SDM is to communicate accurate information about each person’s potential to benefit from screening, especially if it meaningfully differs from the average summary results reported in trials. This is particularly important for LCS. Prior work examining the range of absolute benefits across all individuals enrolled in the National Lung Screening Trial demonstrated that conveying average population information can dramatically overstate or understate lung cancer mortality reduction in thousands of individuals [4]. This is because both lung cancer risk and life expectancy varied substantially across eligible patients, such that the average mortality benefit identified in the National Lung Screening Trial was driven upward by those at the highest risk, whereas many patients experienced a benefit that was far below the average [5,6]. Using prediction-based approaches to estimate individualized net benefits can support the communication of much more accurate information across individuals in a heterogeneous group of screening eligible individuals [7]. Such approaches form an inherently more patient-centered basis for SDM [8].

Objective

However, web-based decision tools that enable quick calculations and intuitive data presentations are needed to streamline the process of obtaining individualized information and more accurately inform patient–clinician LCS discussions in routine practice [9]. However, implementing clinical decision support tools in routine clinical practice has been difficult to achieve. Patient-facing tools have shown promise in improving patients’ understanding of the criteria and procedural requirements for lung screening [10]; however, discussing the details of individualized risks and benefits with patients—a critical aspect of SDM—can be challenging for providers.

Numerous barriers, including infrastructure limitations and clinicians’ perceptions of SDM taking too much time, have led to a lack of uptake in the integration of decision support interventions into standard clinical practice [11,12]. Therefore, the objective of our study is to test two strategies for implementing a prediction-based SDM tool for LCS (DecisionPrecision) [13] in primary care at eight Veterans Affairs (VA) medical centers: (1) a quality improvement (QI) training approach and (2) academic detailing (AD).

Methods

Overview

Our implementation efforts and evaluation of each took place in 2 phases. In phase 1, we used QI training as a strategy for implementing DecisionPrecision as part of a hybrid type 3 implementation study design [14]. Specifically, we used a multisite, cluster-based randomization trial to compare the effectiveness of standard implementation versus the effectiveness of standard implementation plus the Learn, Engage, Act, and Process (LEAP) QI training program [15]. The standard implementation comprised integrating a link to the tool into the VA computerized patient record system (CPRS) and providing educational materials on the tool to a local LCS champion.

Although we observed a substantial number of tool uses, primarily by dedicated screening coordinators (as described in the following sections), results from phase 1 suggested that neither LEAP nor standard implementation contributed to the absolute number of tool uses at a site by primary care providers (PCPs). Consequently, phase 2 of the study switched to a different implementation strategy—namely, AD—and the study design transitioned to a hybrid type 2 implementation study, which comprised coprimary aims: (1) to determine the effectiveness of the clinical intervention (ie, DecisionPrecision) on key outcomes and (2) to determine the potential effectiveness of AD as an implementation strategy for the intervention. In phase 2, the focus of this study was on the second coprimary aim. Findings from our evaluation of the effectiveness of the tool in the LCS decision (first coprimary aim) will be described elsewhere.

The primary outcome measure for both phases is the use of DecisionPrecision at the site level (ie, the absolute number of tool uses at a site over a specific period). This was the best measure of the reach of our implementation strategies that was feasible, given the constraints we faced: use of a standalone tool

https://humanfactors.jmir.org/2022/2/e32399

JMIR Hum Factors 2022 | vol. 9 | iss. 2 | e32399 | p.78
(page number not for citation purposes)
that was deidentified and not connected to the health record and did not allow tracking of which clinicians and patients were associated with specific tool uses. This more ecologic site-level measure does not precisely fit the definition of reach in the reach, effectiveness, adoption, implementation, and maintenance (RE-AIM) framework [16], which defines reach as “the absolute number, proportion, and representativeness of individuals who are willing to participate in a given initiative, intervention, or program.” Moreover, although providers are the ones who must decide if they are willing to use the tool, it is the patients who we are ultimately trying to influence with tool use. Thus, we felt that it was more important to measure the number of tool uses rather than the number of providers using the tool. If 1 provider used the tool once and another provider used it 50 times, we were interested in the fact that the tool was used on 51 patients rather than the fact that it was used by 2 providers. A separate paper will focus on the effect of tool use on patient uptake of LCS (effectiveness from the RE-AIM framework).

In accordance with requirements of the journal, the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and online Telehealth) checklist for reporting the study’s methods and findings was completed (Multimedia Appendix 1).

Common Methodology Across Phase 1 and Phase 2

Setting and Site Selection

We recruited 8 VA sites that participated in the Evaluation of the VA LCS Clinical Demonstration Project (LDCT demo) [17]. These sites were chosen as (1) they had an LCS program currently in place, and (2) as part of the LDCT demo, they used a standard set of clinical reminders built by the VA’s National Center for Health Promotion and Disease Prevention to identify patients eligible for LCS. Clinical reminders are embedded within the electronic health record of the Veterans Health Affairs (CPRS) and can prompt staff to screen patients, review or assess risk factors, or offer interventions and screenings that may be due for an individual patient. The use of these standardized reminders facilitates data collection on LCS eligibility, discussions, and decisions. These advantages provided an ideal setting for testing different implementation strategies for a decision support intervention.

Of these 8 sites, 7 (87%) agreed to participate in our implementation initiative, and 1 (13%) declined because of a competing program. To replace the eighth site, we used VA administrative data to identify sites that used the LCS clinical reminder in CPRS as an indicator of an active LCS program. One of these sites agreed to participate in the study.

Participants

To implement DecisionPrecision, we worked with clinical leaders from each site’s LCS program and screening coordinators when available. These core site team members enlisted others at their sites to help with implementation, including primary care leadership; PCPs; and team members from pulmonology, radiology, and oncology departments. Information technology and data security staff members were also engaged.

Intervention: DecisionPrecision

To meet the need for facilitating SDM and providing individualized information on LCS, we created DecisionPrecision, a provider-facing web-based decision support tool [13]. Our goal was to facilitate individualized and patient-centered SDM within the confines of a busy clinic environment. In particular, the tool seeks to quickly help PCPs tailor the LCS discussion based on the patient’s risk factor profile, more strongly encouraging screening for those with higher predicted lung cancer risk and potentially larger health gains with LCS. To this end, the tool provides the following: (1) individualized quantitative risk assessment of screening trade-offs, aided by an externally validated and accurate lung cancer risk prediction model [18], along with recommendation categories that clarify when screening is and is not preference sensitive [7]; (2) patient-friendly language for the provider to use with the patient; (3) patient-facing graphics, selected based on their ability to help patients understand personalized risks and benefits in a randomized survey experiment [19]; and (4) quick and easy documentation of personalized SDM after using the tool.

The final version of the DecisionPrecision tool tested in this implementation study was the result of multiple rounds of iterative changes that incorporated lessons from an observational field study of patient-clinician conversations in the absence of a decision support tool and iterative feedback and usability testing on multiple tool prototypes from decision aid researchers, PCPs, pulmonologists, screening coordinators, and patients. After phase 2, the tool also underwent extensive updates based on experiences and observations throughout the implementation project and additional feedback from clinicians, screening coordinators, and leadership. Screenshots of DecisionPrecision can be found in Multimedia Appendix 2.

Ethics Approval

This project was categorized as QI and was, therefore, not formally reviewed by the institutional review boards of participating sites [20].

Phase 1: Evaluation of a QI Implementation Strategy

Phase 1 Methods

In phase 1, the 8 participating sites were randomized to either standard (S1, S2, S3, and S4) or enhanced implementation (ie, the LEAP QI training program: E1, E2, E3, and E4), stratified by the rate of clinician completion of the initial LCS clinical reminder (high vs low completion for patients eligible for LCS between May 2015 and June 2015). A CONSORT ( Consolidated Standards of Reporting Trials) diagram summarizing the randomization of the sites is included in Multimedia Appendix 3. The characteristics of the randomized sites are included in Multimedia Appendix 4. Both standard and enhanced implementation strategies are described in detail in the following sections.
**Standard Implementation**

The standard implementation involved the following: (1) adding a tool link to the VA's electronic health record CPRS and (2) promoting the tool via emails, conference calls, and meetings.

**DecisionPrecision Access Within the Electronic Health Record**

A link to the decision tool was embedded within the clinical reminders for PCPs at all sites between August 29, 2017, and October 4, 2017. The language associated with the link and the specific location of the link were established based on conversations with the site team to best fit the mechanisms for the LCS at the site.

**Promotion of DecisionPrecision**

All sites were asked to notify relevant providers about DecisionPrecision. Site leads were given a draft email to providers that could be easily tailored with site-specific information. This email included a brief description of the key features of the tool and a link to a brief YouTube video that describes the tool’s development (eg, how the algorithm was designed) and how to use the website. Educational materials on the tool were provided, including a sample risk assessment from the decision tool, a 1-page attachment explaining how to routinely and quickly use DecisionPrecision to personalize the decision tool, a 1-page attachment explaining how to use the website. Educational materials on the tool were provided, including a sample risk assessment from the decision tool, a 1-page attachment explaining how to routinely and quickly use DecisionPrecision to personalize discussions about LCS, and a screenshot of the link in the CPRS clinical reminder. The site teams were asked to promote decision aid through key local leaders and staff meetings.

**Enhanced Implementation**

Enhanced implementation included all the components available in the standard implementation as well as QI training using the LEAP program. LEAP is a 26-week QI training program to engage frontline clinical teams in using a hands-on learning approach (see Multimedia Appendix 5 for a brief description of the LEAP curriculum). The core components include a structured curriculum that focuses on QI methods, a coaching and learning community, and a web-based platform for sharing videos and other resources. Training goals include gaining confidence in applying QI methods to improve the quality of care within the demands of everyday clinical practice and completing a QI project using plan-do-study-act principles.

Of the 4 sites randomized to LEAP, 3 (75%) participated from February to July 2017 (E1, E2, and E4), and 1 (25%) declined (E3) because of time constraints for the site leads. The participating sites established an interdisciplinary LEAP team comprising 2 to 10 participants. The teams developed and executed a project charter to complete a plan-do-study-act cycle of change related to enhancing SDM for LCS using DecisionPrecision. The team members participating in LEAP were provided with early access to a link to DecisionPrecision in March 2017 so they could access the site as part of their improvement projects.

The improvement project at 1 site (E1) was the development of a process within 1 patient-aligned care team, whereby DecisionPrecision was used to identify eligible patients at the highest risk of lung cancer and then inform the SDM conversation for at least one veteran each week during the LEAP improvement program. The project for the other 2 sites (E2 and E4), which had more centralized screening programs and full-time screening coordinators, was for the screening coordinator to test the use of DecisionPrecision with all new consults for LCS.

**Evaluation Methods**

Phase 1 used a hybrid trial type 3 design [14]. The primary purpose of this design was to evaluate the effectiveness of the implementation strategy, and our primary question was, Does enhanced implementation work better than standard implementation for facilitating the use of DecisionPrecision? Thus, the primary outcome measure for evaluating the effectiveness of the enhanced implementation strategies was the use of DecisionPrecision at each participating site. The absolute tool use data by site were obtained from the DecisionPrecision website. In March 2018, a dropdown box was added asking the provider to indicate their site. To identify sites before March 2018, we used data on IP addresses collected by the website for each record entered and linked each IP address to a study site based on validation against the site data and IP addresses collected after March 2018. The analysis of these data included descriptive statistics of tool use by site. We also conducted brief interviews to assess the clinicians’ impressions of the implementation strategy.

**Phase 1 Results**

As noted under the Setting and Site Selection section, sites were selected for this study based on their participation in the VA’s LCS Clinical Demonstration Project to ensure that all sites had similar LCS programs. However, shortly after the start of the study, we observed that the sites had made some changes to their screening programs. Specifically, some sites had stopped the collection of data required to calculate lung cancer risk (specifically, detailed smoking histories and key data used by DecisionPrecision), some sites stopped the routine use of clinical reminders, some sites no longer used a screening coordinator for conducting SDM, and some sites no longer had a screening coordinator for performing any LCS tasks. As each of these changes had the potential to affect the use of DecisionPrecision, these changes across sites are summarized in Table 1.

Consequently, our findings for this phase are presented by comparing the standard implementation sites to three different groups of facilities: (1) the original 4 facilities randomized to enhanced implementation (intention to treat), (2) the 3 facilities that participated in the enhanced implementation program (as treated), and (3) the 3 facilities that had a full-time screening coordinator engaged in SDM discussions with patients (key resource scenario).
Table 1. Summary of important changes occurring across the study sites after randomization.

<table>
<thead>
<tr>
<th>Changes</th>
<th>Sites randomized to standard implementation</th>
<th>Sites randomized to enhanced implementation with LEAPa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S1</td>
<td>S2</td>
</tr>
<tr>
<td>Lung cancer screening clinical reminders (for providers)?</td>
<td>Limited</td>
<td>Yes</td>
</tr>
<tr>
<td>Screening coordinator for conducting shared decision-making?</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Pack year reminder?</td>
<td>Limited</td>
<td>Yes</td>
</tr>
</tbody>
</table>

bAfter randomization, this site decided not to participate in LEAP but continued to participate in the overall trial.

table 2 shows the data on tool use for all sites in the 6 months following the time at which they all had access to DecisionPrecision but before phase 2 (AD) was initiated in April 2018. It also indicates the sites in each comparison group.

Table 2. Tool use each month by site (number of patients).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening coordinator</td>
<td>E1</td>
<td>7</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>23</td>
<td>3.8 (2.1)</td>
</tr>
<tr>
<td></td>
<td>E2</td>
<td>54</td>
<td>80</td>
<td>52</td>
<td>74</td>
<td>72</td>
<td>85</td>
<td>417</td>
<td>69.5 (13.6)</td>
</tr>
<tr>
<td></td>
<td>E4</td>
<td>42</td>
<td>47</td>
<td>40</td>
<td>40</td>
<td>30</td>
<td>34</td>
<td>233</td>
<td>38.8 (6.0)</td>
</tr>
<tr>
<td></td>
<td>E3</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>33</td>
<td>23</td>
<td>22</td>
<td>88</td>
<td>14.7 (13.2)</td>
</tr>
<tr>
<td></td>
<td>S1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0.3 (N/A)</td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.2 (N/A)</td>
</tr>
<tr>
<td></td>
<td>S3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0.3 (N/A)</td>
</tr>
<tr>
<td></td>
<td>S4</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>9</td>
<td>1.5 (0.8)</td>
</tr>
</tbody>
</table>

aRandomized to enhanced implementation.
cParticipated in LEAP QI training program.
dStaffed by a lung cancer screening coordinator.
eN/A: not applicable.

Table 3 presents the 6-month mean number of tool uses between the standard implementation sites versus the 3 comparison groups. All 3 comparisons showed significantly less tool use for the standard implementation sites. However, the presence of a screening coordinator at 3 of the enhanced implementation sites and none of the standard implementation sites greatly confounded these comparisons.

Table 3. Mean number of tool uses over 6 months: 4 standard implementation sites as compared with 3 ways of grouping intervention sites.

<table>
<thead>
<tr>
<th>Comparisons</th>
<th>Tool uses: 6 months, mean (SD)a</th>
<th>Median difference b</th>
<th>Median difference (95% CI)b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard implementation sites (A) versus the original 4 facilities randomized to enhanced implementation (intention to treat; B1)</td>
<td>3.5 (3.7)</td>
<td>190.3 (174.8)</td>
<td>0.03</td>
</tr>
<tr>
<td>Standard implementation sites (A) versus the 3 facilities that participated in the enhanced implementation program (as treated); B2</td>
<td>3.5 (3.7)</td>
<td>224.3 (197.1)</td>
<td>0.049</td>
</tr>
<tr>
<td>Standard implementation sites (A) versus the 3 facilities that had a full-time screening coordinator engaged in SDM discussions with patients (key resource scenario; B3)</td>
<td>3.5 (3.7)</td>
<td>246.0 (164.9)</td>
<td>0.049</td>
</tr>
</tbody>
</table>

aTotal tool uses across all the sites in the group divided by the number of sites in the group.
bOn the basis of the Wilcoxon rank-sum test using the medians of the differences (Hodges-Lehmann estimator) between sites in the 2 groups.
cSDM: shared decision-making.
Phase 1 Discussion

Our phase 1 findings showed that sites that participated in the LEAP QI training program used DecisionPrecision significantly more often than the standard implementation sites. However, there is some indication that participation in LEAP was not the primary contributing reason for these findings, as the implementation arms were imbalanced with regard to the presence of a screening coordinator. As evidence, a site in the enhanced implementation group that did not participate in LEAP (E3) but had a screening coordinator adopted and used the tool. In addition, a site in the enhanced implementation group that did not adopt the tool in routine use (E1) was also the site in the group that no longer used a screening coordinator. Although we could not determine tool use by provider type (PCP vs screening coordinator) from the website data, the vast majority of tool uses at sites E2, E3, and E4 was by screening coordinators based on tool use data collected manually by the coordinators, which showed numbers comparable with those obtained from the website.

Although participation in the LEAP may have contributed to the increased use of the tool, we conclude that the existence of a screening coordinator likely played a much larger role in tool use. Therefore, the question remains whether the screening coordinators who adopted the tool would have used it to a lesser extent if they had not participated in LEAP. Although the site with a coordinator who did not participate in LEAP showed the lowest tool use of the 3 sites with coordinators, other data (not shown) showing tool use as a percentage of all eligible patients indicated that E3 had a comparable rate of tool use with that of E2 and E4.

In addition, feedback from one of the screening coordinators suggested that QI training was not key to implementing the tool:

Well, it [LEAP] seemed to be more tied towards quality process improvement, so it was helpful for that. When it comes to the decision precision tool, I’m not really sure if I can concretely tie it to that.

QI collaboratives, including internet-based videoconferencing adaptations, are a common approach to helping health care teams implement new initiatives or improve existing programs [21]. However, evidence for their success is mixed, including weaknesses in the reporting of methods and potential publication bias. Nevertheless, findings from several studies have shed light on some factors that are correlated with successful implementation. A study of 11 collaboratives focusing on 11 different topics found that innovation attributes, organizational support, innovative team culture, and professionals’ commitment to change are instrumental to perceived effectiveness [22]. With specific regard to an innovation’s attributes, the study found that the newer working methods were perceived by professionals as having relative benefit, being compatible with norms and values, not difficult to learn and implement, and leading to observable results, the more the implementation process was perceived as successful. This finding is certainly consistent with feedback from screening coordinators, who all perceived DecisionPrecision as doing an excellent job in conveying important information on risks and benefits to patients. They found that the tool was relatively easy to use and incorporate into their workflow.

A systematic review of QI collaboratives concluded that collaboratives reporting success generally addressed relatively straightforward aspects of care and had a strong evidence base [21]. The implementation of DecisionPrecision could be considered straightforward in that it required only the addition of a link in the electronic medical record and did not require any significant changes in procedures or workflows. In addition, the scientific evidence underlying a prediction-based approach to LCS is relatively strong.

Furthermore, findings from a study on the effect of a learning collaborative on colorectal cancer screening rates in primary care practices are also consistent with ours [23]. Specifically, the teams had difficulty spreading the change beyond the clinicians who participated in the collaboration:

Other clinicians in a practice tended not to be aware of or engaged in the CRC (colorectal cancer) improvement efforts, and teams tended to communicate poorly with the rest of the practice regarding QI plans.

As a result—as occurred in our case—other providers, most notably PCPs, were not engaged and did not adopt the tool.

Another similarity between our study and the colorectal cancer screening study was that the clinicians who participated in the collaboratives (screening coordinators in our case) were very motivated to use the tool to improve the LCS process, which may not be the case among clinicians who did not participate (eg, PCPs).

Whether LEAP had a significant impact on the absolute number of DecisionPrecision uses by the screening coordinators, the collaborative approach did not have the intended effect of engaging the broader community of PCPs in using the tool. Feedback from participants in LEAP suggested the potential utility of an alternative strategy, namely one that focuses on one-on-one conversations with clinicians about the tool. Of the 26 sessions of the LEAP program, 1 (4%) was devoted to presenting and discussing the evidence behind the tool and how to use it. One screening coordinator noted the following:

I think the only helpful parts of it [LEAP], when it came to trying to implement the DecisionPrecision tool, was talking with the team...about what the stratified risks really mean...how you can come up with things like personalized harms and having that shared decision-making conversations where things are more preference-based—understanding that piece was extremely helpful and I can say that now, hindsight being 20/20 and having done a ton of shared decision-making in the last couple of years, I don’t think I could’ve done it as effectively if I didn’t have the knowledge that [was] shared with us during the LEAP program.

On the basis of the phase 1 findings, we switched to a different implementation strategy in phase 2—namely, AD.
AD was selected as an implementation strategy to convey information directly and one on one to PCPs about the evidence behind prediction-based screening and explain how to use DecisionPrecision. As noted on the National Resource for AD website [24], busy clinicians need an accurate source of current data on the effectiveness of current interventions. However, they have many competing demands for their time. Trying to assemble current evidence from a continuous influx of research is incredibly challenging to do on one’s own. AD combines a one-on-one outreach approach with the best available evidence. We hired and trained a master’s student in public health to meet clinicians to assess individual needs and then offer tailored, evidence-based advice for using DecisionPrecision as part of the LCS SDM process.

Phase 2: Academic Detailing

Phase 2 Methods

Our AD strategy focused on directly engaging PCPs, in addition to screening coordinators. We decided to use this strategy at all 8 participating sites rather than randomize sites to AD versus standard implementation in an effort to conduct a more extensive formative evaluation of the AD process, which, to the best of our knowledge, has not before been used to promote the use of a prediction-based SDM tool.

The goal of AD was to encourage providers, through the use of the DecisionPrecision tool, to adopt a prediction-based approach to tailoring how strongly screening is encouraged (based on estimated net benefit for the individual and consideration of how preference sensitive the decision is) to thereby facilitate a brief everyday SDM discussion and make decision-making more patient centered [7,8]. AD site visits were offered to all sites; of the 8 sites, 7 (87%) agreed to the site detailing visits. One of the sites underwent substantial workforce changes during the study and opted not to participate in the AD. Heads of primary care were asked to send emails to PCPs announcing the detailed visit and the purpose of the visit and encourage providers to participate in one-on-one detailing.

AD materials, which were developed and available for use during meetings, included a 4-page visual abstract of the evidence behind, benefits of, and key features of DecisionPrecision; a pocket card on how to use the tool; a CPRS clinical reminder screenshots and tool link handout; a handout on how to copy a templated description of the SDM discussion into the medical record; a list of references (in case of questions or concerns about the evidence); and a business card that included the URL to the DecisionPrecision website. The key information presented during these meetings were (1) how using a prediction-based approach for LCS can improve quality of care and (2) how to use the DecisionPrecision tool with eligible veterans to inform more patient-centered SDM and tailor screening encouragement during SDM discussions. At the end of each detailing session, the academic detailer asked for a provider’s commitment to using the decision tool in the next 1 to 2 weeks and for permission to follow up with them 3 to 4 weeks after the detailing visit [25]. See Multimedia Appendix 6 for a summary of the characteristics of the AD strategy per the published recommendations.

We conducted semistructured phone interviews with a sample of PCPs for 2 to 4 weeks following their AD visits. The interviews included questions on the utility of the AD visit, use of DecisionPrecision since the visit, usefulness of the tool, ease of tool use, challenges in using the tool, and suggestions for improving the tool. Audio-recorded interviews lasted approximately 20 minutes and were transcribed verbatim.

Evaluation Methods

As noted in the Introduction section, this paper presents data on the potential effectiveness of AD as a strategy for promoting the use of DecisionPrecision. Data on the effectiveness of the tool as a clinical intervention for improving the quality of LCS decisions have been presented elsewhere. The effectiveness of AD as an implementation strategy was assessed by (1) examining tool use following the AD visits and (2) conducting semistructured interviews with a subset of PCPs following their participation in AD visits.

Analysis of Tool Use

We conducted an interrupted time-series analysis to determine whether there was a difference in the overall tool use between the 6 months following the initiation of AD and the 6 months following the initiation of enhanced implementation (and before AD). We fitted a linear mixed model with the study period as the fixed effect of interest and a random intercept for each site.

Post-AD Interviews

We used NVivo (version 12; QSR International) to conduct an inductive thematic content analysis of the postdetailing interviews, searching for themes that emerged from the qualitative data. Team discussion of the findings led to agreement on the common themes, which included the major barriers to tool use and the features of the tool that the providers found to be beneficial.

Phase 2 Results

We examined 105 PCPs at the 7 participating sites from June to October 2018 (E3 chose not to participate). Each site visit lasted 2 to 3 days, except for E1, where visits occurred over 2 months. The academic detailer met providers in primary care clinics, primary care resident clinics, and community-based outpatient clinics. Snowball sampling was used to identify providers before and during the site visit. Individual meetings were tailored to provider needs in terms of both content and duration. The duration of the meetings ranged from 4 to 40 minutes, with a mean duration of 13 minutes. Most meetings were one on one; however, a few meetings were with 2 to 3 providers simultaneously.

Tool Use by Site

Table 4 shows data on absolute tool use for the 6 months following the initiation of AD for all sites participating in AD. For comparison purposes, the last 2 columns of the table also show the total and average tool use for a similar period (6 months) before the AD intervention.
### Table 4. Monthly tool use at seven sites participating in AD<sup>a,b</sup> (number of patients).

<table>
<thead>
<tr>
<th>Sites</th>
<th>Before AD</th>
<th>Total 6 months after AD&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Per month after AD, mean (SD)</th>
<th>Total 6 months before AD&lt;sup&gt;d&lt;/sup&gt;</th>
<th>Per month before AD, mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All sites</td>
<td>84</td>
<td>90</td>
<td>114</td>
<td>137</td>
<td>137</td>
</tr>
<tr>
<td>E1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>E2</td>
<td>52</td>
<td>62</td>
<td>65</td>
<td>72</td>
<td>73</td>
</tr>
<tr>
<td>E4</td>
<td>20</td>
<td>20</td>
<td>45</td>
<td>51</td>
<td>51</td>
</tr>
<tr>
<td>S1</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>S2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>S3</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>S4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

<sup>a</sup>AD: academic detailing.

<sup>b</sup>The site designations, Enhanced Implementation and Standard Implementation, are not relevant for phase 2 as all sites received academic detailing; however, the labeling was maintained for linking to the phase 1 data (last 2 columns).

<sup>c</sup>Pre-AD months: October 2017 to March 2018; data are pulled from Table 2.

<sup>d</sup>N/A: not available.

An interrupted time-series analysis showed no significant difference in tool use between pre- and post-AD periods (95% CI 5.06-6.40, fewer tool uses after AD to more tool uses after AD; P=.82). Thus, it appears that the introduction of AD as an implementation strategy did not encourage substantial additional tool use (see Multimedia Appendix 7 for detailed results of this analysis).

**Interview Responses**

Of the 105 PCPs who participated in the AD, 83 (79%) provided their contact information for follow-up purposes. Of the 83 providers contacted, 33 (40%) participated in the post-AD follow-up interviews. Virtually unanimously, the participants felt that the AD visit helped provide them with an explanation of the tool’s purpose, the science underlying its development, and how to use it. Respondents appreciated the concise presentation and opportunity for a one-on-one discussion to walk them through the tool. Regarding the tool itself, many saw value in the tool’s ability to (1) shape clinician feelings about the LCS and (2) guide a useful approach to LCS discussions. They also felt that the tool enhanced their ability to share information about individualized risks and the pros and cons of screening. However, at follow-up, a few PCPs (6/33, 18%) used the tool with an actual patient. Limited time in the clinic was perceived as a key barrier by almost all the PCPs. Most PCPs reported needing 1 to 2 minutes to discuss LCS but frequently voiced not having even 1 to 2 minutes during a visit because of patient-specific needs that were a higher priority (eg, acute complaints) or organizational priorities (eg, performance measures). Similarly, having to input clinical data on risk factors into the tool was seen as a significant barrier to tool use as it added more time to the visit.

**Phase 2 Discussion**

In phase 2, there was no difference in tool use between before and after AD. Follow-up interviews with PCPs indicated that the AD strategy increased provider awareness and appreciation for the benefits of the tool. However, other priorities and limited time prevented PCPs from using them during routine clinical visits.

We decided to pursue AD as a strategy because of the consistent literature documenting the effectiveness of this approach for aligning clinician behavior with evidence-based best practices [26-29]. Others have emphasized that for eHealth to optimize preventive care, electronic risk factor data need to be seen as relevant and useful by PCPs [30]. Before our AD intervention, we had limited success in engaging PCPs in considering the important role of overall lung cancer when making LCS decisions. We needed a strategy that could help us have meaningful conversations with frontline clinicians making daily decisions about LCS.

Although rooted in a strong evidence base, prediction-based approaches to decision-making about cancer screening are relatively novel and unfamiliar to many clinicians [31,32]. There are potential cognitive challenges with moving from SDM that conveys population average information to a one-size-fits-all approach for all patients who meet eligibility criteria and toward a prediction-based approach that tailors the strength of the recommendation based on the degree of estimated net benefit for each eligible individual. Given the frequency of cancer screening decisions in primary care and how entrenched practice styles and decision-making can be for such practices, it was unclear whether a brief AD intervention would be able to successfully convey the rationale for a prediction-based approach. Most AD studies have focused on medication prescribing, and few such studies used detailing to modify a...
A decision-making approach to a commonly delivered category of service such as cancer screening may involve numerous steps, factors, and challenges. On the basis of feedback from follow-up interviews with participating providers, AD allowed us to meaningfully engage dozens of PCPs across 7 sites. At the very least, these providers are now familiar with the prediction-based approach to SDM for LCS, the tool, and how it can be used in a busy primary care setting. Interview responses suggest that we have changed the way some clinicians think about decision-making for LCS, especially their understanding of the utility of a prediction-based approach to screening decisions. Nevertheless, our light-touch AD did not result in the routine use of the tool among PCPs in our study sample, primarily because of time constraints. As others have noted, the incentives (and disincentives) in our health care system will need to change if providers are to have sufficient time to engage in SDM [33].

General Discussion

Limitations

The limitations of this study include the small sample size (8 sites). Other studies have emphasized the importance of examining whether specific components of multicomponent implementation strategies have stronger associations with absolute tool use than other components in an effort to streamline these strategies to be more time and cost-effective [34]. A larger sample size would have enabled us to examine whether specific components of LEAP were associated with the use of DecisionPrecision. An additional limitation was that enhanced implementation was confounded by the presence of a screening coordinator, making it impossible to attribute the findings to the enhanced implementation strategy alone. In phase 2, the AD intervention was based on a single site visit and a single one-on-one conversation with PCPs rather than multiple reinforcing visits or a longitudinal relationship with PCPs.

Conclusions

The phase 1 findings do not provide conclusive evidence of the benefit of a QI training approach for implementing a decision support tool for LCS among PCPs. Screening coordinators in the study used the tool frequently, and it is possible that the LEAP program helped them adopt the tool. However, other factors may have contributed to tool use, including the coordinators’ perception of the added benefit of using the tool as part of their responsibility for educating patients on LCS and the relative ease of incorporating tool use into their workflow. As PCPs were not engaged in the phase 1 implementation strategy, the phase 2 implementation strategy—AD—targeted these clinicians. On the basis of our experience with phase 1, the focus of the AD approach was to educate PCPs on the benefits of tool use and discuss the best ways of incorporating it into their clinical practice. However, even when PCPs see value in a prediction-based approach to LCS decision-making and a tool to support that approach, they face major challenges in implementing it in a busy primary care clinic. This was a consistent finding across all study sites. Thus, in terms of the RE-AIM framework, we feel that the adoption (willingness to initiate a program) and reach (willingness to use) of our standalone DecisionPrecision tool, if left unchanged, is likely to be limited among PCPs within other health care settings as well.

One implication of these findings for implementing decision support tools for LCS—and potentially other cancer screening tools—is that QI as an implementation strategy may not be helpful; instead, the focus of implementation should be on working with individual clinicians and screening coordinators to promote tool use. Screening coordinators bought into the rationale for using the tool and were able to adopt and use it routinely (high awareness and good reach among coordinators). However, our light-touch, single-visit AD strategy did not affect tool use among PCPs, although feedback from PCPs suggested that this strategy did achieve our goal of increasing provider awareness and appreciation of the benefits of the tool (awareness but limited adoption and reach among PCPs). Other priorities and limited time prevented PCPs from using them during routine primary care clinic visits. These barriers point to the second implication of our findings; namely, prediction-based SDM tools need to be automated as much as possible for use in primary care to better integrate into workflows and help PCPs more quickly understand how to prioritize LCS discussions among other competing demands [35]. Regarding the latter need, an ongoing Agency for Health Research and Quality–funded project is addressing this barrier by automating predictions and integrating the DecisionPrecision tool within multiple electronic health records, including Epic, Cerner, and CPRS health record systems. CPRS is still used in most VA health systems at present before a planned transition to Cerner [36-39].

Acknowledgments

TJC is supported by a Career Development Award from Veterans Affairs Health Services Research and Development (CDA 16-151). This study was conducted with a grant from the Veterans Affairs Quality Enhancement Research Initiative Program (QUE 15-286).

Conflicts of Interest

At the time of the study, all the authors were employees of the Department of Veterans Affairs (VA). This work was funded by the Department of VA Health Services Research and Development Quality Enhancement Research Initiative grant (QUE 15-286) and the VA Career Development Award (CDA 16-151). TJC and AF created an invention and submitted an invention disclosure form with the VA for the web-based lung cancer screening (LCS) risk calculator, screenLC (previously known as, and referenced throughout the manuscript, as DecisionPrecision) [19], but did not receive financial support for this invention. RSW participates
as a panelist and coauthors several guidelines or papers related to the implementation of LCS and serves as a co-chair of the VISN1 LCS implementation task force but does not receive dedicated salary support for this role. JL is the chair of VA's Scientific Merit Review Board, which awards VA Health Services Research and Development grant funds. This organization funded the project on which we are reporting but was not involved in any way in the review and discussions regarding this grant. Authors TJC, AF, RSW, and JL also coauthored a related manuscript [40].

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 658 KB - humanfactors_v9i2e32399_app1.pdf]

Multimedia Appendix 2
DecisionPrecision screenshots.
[DOCX File, 340 KB - humanfactors_v9i2e32399_app2.docx]

Multimedia Appendix 3
CONSORT (Consolidated Standards of Reporting Trials) diagram, phase 1.
[DOC File, 32 KB - humanfactors_v9i2e32399_app3.doc]

Multimedia Appendix 4
Site characteristics.
[DOCX File, 17 KB - humanfactors_v9i2e32399_app4.docx]

Multimedia Appendix 5
Learn, Engage, Act, and Process (LEAP) curriculum.
[PNG File, 213 KB - humanfactors_v9i2e32399_app5.png]

Multimedia Appendix 6
Characteristics of the academic detailing strategy.
[DOCX File, 15 KB - humanfactors_v9i2e32399_app6.docx]

Multimedia Appendix 7
Results of the interrupted time-series analysis.
[DOCX File, 14 KB - humanfactors_v9i2e32399_app7.docx]

References


Abbreviations

AD: academic detailing

CONSORT: Consolidated Standards of Reporting Trials

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and online Telehealth

CPRS: computerized patient record system

LCS: lung cancer screening

LDCT: low-dose computed tomography

LEAP: Learn, Engage, Act, and Process

PCP: primary care physician

QI: quality improvement

RE-AIM: reach, effectiveness, adoption, implementation, and maintenance

SDM: shared decision-making

VA: Veterans Affairs
©Julie Lowery, Angela Fagerlin, Angela R Larkin, Renda S Wiener, Sarah E Skurla, Tanner J Caverly. Originally published in JMI Human Factors (https://humanfactors.jmir.org), 01.04.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMI Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Requirements for a Bespoke Intensive Care Unit Dashboard in Response to the COVID-19 Pandemic: Semistructured Interview Study

Brittany Davidson1,2, MA, PhD; Katiuska Mara Ferrer Portillo3, LLB, MA, MSc; Marceli Wac1,4, BSc; Chris McWilliams1, BSc, MSc, PhD; Christopher Bourdeaux4, BCHIR, MA, MB; Ian Craddock1, BEng, PhD

1Department of Electrical & Electronic Engineering, University of Bristol, Bristol, United Kingdom
2School of Management, University of Bath, Bath, United Kingdom
3French Department, School of Modern Languages, University of Bristol, Bristol, United Kingdom
4University Hospitals Bristol and Weston National Health Service Foundation Trust, Bristol, United Kingdom

Corresponding Author:
Marceli Wac, BSc
Department of Electrical & Electronic Engineering
University of Bristol
1 Cathedral Square
Bristol, BS1 5DD
United Kingdom
Phone: 44 07774286342
Email: m.wac@bristol.ac.uk

Abstract

Background: Intensive care units (ICUs) around the world are in high demand due to patients with COVID-19 requiring hospitalization. As researchers at the University of Bristol, we were approached to develop a bespoke data visualization dashboard to assist two local ICUs during the pandemic that will centralize disparate data sources in the ICU to help reduce the cognitive load on busy ICU staff in the ever-evolving pandemic.

Objective: The aim of this study was to conduct interviews with ICU staff in University Hospitals Bristol and Weston National Health Service Foundation Trust to elicit requirements for a bespoke dashboard to monitor the high volume of patients, particularly during the COVID-19 pandemic.

Methods: We conducted six semistructured interviews with clinical staff to obtain an overview of their requirements for the dashboard and to ensure its ultimate suitability for end users. Interview questions aimed to understand the job roles undertaken in the ICU, potential uses of the dashboard, specific issues associated with managing COVID-19 patients, key data of interest, and any concerns about the introduction of a dashboard into the ICU.

Results: From our interviews, we found the following design requirements: (1) a flexible dashboard, where the functionality can be updated quickly and effectively to respond to emerging information about the management of this new disease; (2) a mobile dashboard, which allows staff to move around on wards with a dashboard, thus potentially replacing paper forms to enable detailed and consistent data entry; (3) a customizable and intuitive dashboard, where individual users would be able to customize the appearance of the dashboard to suit their role; (4) real-time data and trend analysis via informative data visualizations that help busy ICU staff to understand a patient’s clinical trajectory; and (5) the ability to manage tasks and staff, tracking both staff and patient movements, handovers, and task monitoring to ensure the highest quality of care.

Conclusions: The findings of this study confirm that digital solutions for ICU use would potentially reduce the cognitive load of ICU staff and reduce clinical errors at a time of notably high demand of intensive health care.

(JMIR Hum Factors 2022;9(2):e30523) doi:10.2196/30523

KEYWORDS
intensive care; critical care; COVID-19; human-centered design; dashboard; eHealth; disease monitoring; monitoring; ICU; design; development; interview
**Introduction**

**Background**

The intensive care unit (ICU) is a busy working environment where a variety of clinical staff perform different duties at scheduled times of the day, while also having to respond to unexpected, often critical issues with patients. ICUs are typically heavily instrumented, and staff need to be alert to many sources of data from equipment such as ventilators as well as the patients’ vital signs and lab test results. For a complex ICU patient (eg, those with multiple conditions), anywhere between 80 and 200 medical interventions are delivered daily and, prior to the COVID-19 pandemic, a member of ICU staff would typically be responsible for up to 10 patients each day [1]. Nonoptimal decisions and clinical errors in this cognitively demanding environment are known to impact patient outcomes [1-4], and a large body of evidence demonstrates that working in an ICU is highly stressful [5,6].

Much of the relevant data for clinical decision-making is already available to staff in the ICU. However, this information is typically scattered across a number of applications, devices, and pieces of paper within the ward. Hence, ICU staff may inadvertently fail to notice signs of a patient’s deterioration and struggle to effectively communicate patient updates (eg, test results, medication) or patient requirements (eg, changing tubes, sedative drug management), which will contribute to worse patient outcomes [3,7]. Additional problems such as equipment failures [3] further add to the complexity of working in the ICU and the importance of clear communication among ICU staff [3,8].

The COVID-19 pandemic has generated unprecedented challenges around the globe [9,10], with particularly detrimental impacts on health care systems [10,11]. Increased hospitalizations from COVID-19 put an additional strain on ICU resources, specifically beds with mechanical ventilation [11,12]. In the United Kingdom, this shortage has been such a concern that additional intensive care capacity was made available by the construction of 11 temporary “Nightingale” hospitals [13]. As the pandemic grew in early 2020, the two local ICUs in University Hospitals Bristol and Weston National Health Service Foundation Trust reported a critical need for an information technology solution to help their staff manage ICU dashboards have been used in Brazil to allocate resources and to obtain near real-time information on suspected and confirmed COVID-19 cases [21]; however, we are not aware of any published study that has interviewed clinical staff and broadly captured the requirements for such a dashboard during the pandemic. Such a study will therefore provide both COVID-19–specific and extreme environment–specific working practices and insights [22]. With new highly transmittable COVID-19 variants emerging and with no guarantee that vaccines will engender an immune response against all future variants, it is hoped that the methodology and the findings in this contribution will be of value to those in a position to develop much-needed new technologies for the clinical frontline.

When building tools and devices, it is important to take a human-centered approach to the design of clinical technologies by including the end-user needs as early as possible [14,23,24]. For reasons of efficiency, patient safety, and job satisfaction, clinical professionals should have a direct role in the design of the tools they will use for their jobs [23-26]. Further, this helps to ensure that critical features in design, particularly in terms of visual displays, are not misunderstood [17]. Therefore, this approach helps to ensure that such tools are developed as a result of clinical “pull,” not technological “push,” and helps to prevent technology resistance or avoidance [24,27]. Hence, our approach was to commence interviews with clinical staff within the two local ICUs to elicit their requirements for a clinical dashboard. Due to the extreme pressure on staff during the pandemic, the contagious nature of the disease, and the considerable time pressure to implement a solution, these interviews were conducted remotely with only a minimum sample of available...
staff with various responsibilities to capture a wide range of requirements.

Since hospitals were dealing with a high demand for ICU beds [28], the aim of this study was to obtain important information to help us understand what design aspects of this tool could help to reduce the workload of clinical staff [4]. Such information could be used to design better tools to help with clinical decision-making processes [4]. Similarly, this may reveal a better understanding of which tasks performed by ICU staff may benefit the most from the introduction of such a dashboard. Further, we anticipate that understanding the various responsibilities of ICU staff could be a useful guide for future participatory design work focused on specific health care staff functions. Therefore, taking into account the preferences of end users for the dashboard design, including how data are displayed (eg, table, figure, graph), enables gaining a better understanding of the role for implementation of these preferences in a multifunctional and customizable dashboard, which could help to prepare ICUs for any future waves and mutations of the virus [29,30]. However, this may require further innovations in clinical processes. We are aware of the complexities of health information systems; therefore, the development and integration of new systems require careful consideration and human-centered design [23]. If this is not the case, there is a risk of developing new systems surplus to requirements, or causing further complications if staff need to switch between multiple systems. New systems could potentially compromise patient safety, be difficult to use and learn, and encounter resistance from staff, potentially resulting in poor uptake of the system [23,26].

Hence, the overall aim of this project was to collate a series of insights and requirements from end users across two ICUs to design and build a clinical dashboard to support their increased workload during the COVID-19 pandemic. Requirements elicitation, as seen in this work, by its nature is difficult, as requirements are volatile and necessitate translation from natural language via stakeholders through to tangible software [31,32], especially with a fairly unknown disease that continues to change, mutate, and impact society. The current pandemic has been an “extreme environment” [22] for researchers across disciplines. Hence, this work provides interesting insights regarding COVID-19 specifically but also provides an example of work conducted “in the wild” that encompasses the context, nuances, and uncertainties faced by both the researchers and the ICU staff [33]. The difficulties range from the time and additional pressure ICU staff were under at the time of writing but also entire workforces required to “work from home,” thus diminishing the inability to travel and collaboratively work together to previous expectations. Altogether, this provides a novel frame for this research to occur from initial planning, interviews, through to subsequent testing evaluation, and future iterations being deployed as this work is ongoing.

Methods

Research Questions

At a time of unprecedented workloads in the ICU, clinical staff time was in short supply. We interviewed six staff members working in ICU wards across two hospitals for the UK National Health Service (NHS) (Table 1). We were acutely aware of the additional strain on the NHS and staff; hence, we proceeded with interviews as a direct and simple method to capture requirements and reduce additional workload on ICU staff [31,32].

<table>
<thead>
<tr>
<th>Participant ID number</th>
<th>Job title</th>
<th>Hospital</th>
</tr>
</thead>
<tbody>
<tr>
<td>1099</td>
<td>Consultant in Intensive Care Medicine</td>
<td>A</td>
</tr>
<tr>
<td>1159</td>
<td>Sister (Band 7, Manager)</td>
<td>B</td>
</tr>
<tr>
<td>1252</td>
<td>Anesthetic Registrar</td>
<td>B</td>
</tr>
<tr>
<td>1587</td>
<td>Consultant in Intensive Care Medicine</td>
<td>B</td>
</tr>
<tr>
<td>1704</td>
<td>Consultant in Intensive Care Medicine</td>
<td>B</td>
</tr>
<tr>
<td>1839</td>
<td>Matron (Senior Nurse/Nurse Manager) in Intensive Care</td>
<td>A</td>
</tr>
</tbody>
</table>

Our exploratory interview questions were:

1. What tasks do each participant perform in the ICU, and where would a portable dashboard be helpful to improve safety or reduce workload on staff?
2. Do COVID-19 patients pose particular clinical challenges that need to be taken into account in the dashboard design?
3. Which data would staff need to look at on the dashboard and how should they be presented?
4. Would staff have any concerns about the introduction of a portable dashboard into the ICU?

Participating Hospitals

The two hospitals participating in the study were distinct from one another, with one (hospital A) having a more technologically enhanced ICU where many systems are already digitized in comparison to hospital B, which remains more paper-based with a much larger ICU unit. The interviews were semi-structured, lasting approximately 30-45 minutes, and due to the pandemic situation in early 2020 were conducted exclusively online. Topic guides were used to assist the interviewer and encourage consistency, with additional post hoc questions to further explore any potential themes. The interviews were recorded and transcribed verbatim. We used
Analysis

To analyze these interviews fully, we took a multireasoning approach within our thematic analysis, which consisted of both deductive and inductive approaches to ensure rigorous development of coding trees [36,37]. Initially, we developed a preliminary codebook deductively via the data familiarization phase. We then entered a second phase that took an inductive approach to allow new themes to materialize as we coded each interview. We iterated through this process until the researchers reached a unanimous decision on the final codes, themes, and how they fit together. This in-depth and rigorous methodology aimed to ensure we were as thorough as possible for capturing requirements, alongside taking advantage of our multidisciplinary team, noting that the qualitative coders were not medically trained unlike other members of the team. Hence, it was important for us to ensure that we were accurate and appropriate in our codes, themes, and understanding [38,39]. Therefore, we sought medical insight and advice at every stage of the process within our team. We produced two distinct codebooks: one relating specifically to the requirements of a dashboard and another that related specifically to concerns of dashboard use in ICUs. The requirements codebook consisted of 96 codes that were initially subdivided into the following three requirement categories: 24 technical codes, 56 clinical codes, and 16 operational/logistical codes (see Multimedia Appendix 1). The concerns relating to the dashboard use codebook consisted of 24 codes, including 9 codes regarding design and 15 codes concerning operations (see Multimedia Appendix 2). The final codebooks were then used as the guide to code the qualitative interviews; therefore, in NVivo, each code would have words, phrases, and quotes from participants organized into these codes and themes. This ensured a straightforward and well-organized analysis. Hence, our hybrid approach [37,40] was the method employed for this codification, which identified existing patterns and subsequently regrouping codes into their emerging themes.

Ethics Statement

This work was approved by the Faculty of Engineering Research Ethics Committee at University of Bristol (case 2020-3236).

Results

Overview of Key Requirements

Based on our interviews, we elicited five key requirements from a range of ICU staff to capture a wide range of roles and needs when using clinical dashboards. For instance, a dashboard must be adaptive and flexible to continue to be useful in changing clinical environments. Furthermore, dashboards need to be customizable because different staff may have specific parameters and information they need to see first (eg, “condition at a glance”). This can be achieved by ensuring there is some customizability for individuals or staff groups to select the information that is displayed. A dashboard would need to be mobile, which could reduce reliance on paper-based forms with patients. This also relates to task and patient management, where several staff noted that a dashboard could help with data entry and management alongside assisting with patient handovers, as all information will be collated into one system that is easily accessible. In contrast, some concerns were raised, for instance around infection control when carrying devices in and out of ICUs and high-risk-of-infection areas.

Requirements for an ICU Dashboard

Flexibility With Changing Protocols for an Evolving Disease

Unsurprisingly, the new challenges posed by the COVID-19 pandemic featured heavily in the participants’ responses to our interviews (see Multimedia Appendices 1 and 2). As mentioned above, as these interviews were conducted in early 2020, relatively little was known about the nature of this disease and how to contain or treat it. This had direct implications for hospitals, highlighted by Interviewee 1587 (ICU Consultant), who described ICU staff difficulties handling COVID-19 patients while “keeping on top of constantly changing best practices.” At the time, no consensus had been reached regarding the clinical management protocols for COVID-19 patients, with staff under pressure to run additional tests while keeping track of the results. According to Interviewee 1704 (ICU Consultant), this absence of clear protocols created a very difficult working environment for staff:

...for a couple of days when I was on the COVID-19 side, full witness to it, it was a bit of a shambles and very stressful for the staff. There was no harm done to any of the patients, but it was asking staff to work outside their comfort zone and people found that professionally very difficult...

As ICUs are data-intensive environments by their nature, while in the process of adapting to a new disease and ever-changing protocols, a dashboard could help avoid staff forgetting to check certain parameters or simply avoid them due to being overwhelmed with new tasks and information. For instance, Interviewee 1099 (ICU Consultant) stated: “[a dashboard has the ability to] draw[s] my attention to things that I would probably forget about if I’m honest. I can only process X amount of information.”

One of the great challenges for medics and hospitals during the pandemic has been keeping on top of the new, fast-moving information and updates regarding COVID-19 management and treatment:

We are still learning a lot about [COVID-19 patients]. When [information] was coming from Italy, there was a lot of talk about how the patients were and how we were supposed to treat them, pretty much everything we were told has been wrong. It seems to be a very unusual disease and it’s not like anything we have seen. Whilst we were told it was a really bad pneumonia, which we had to treat with aggressive ventilation; it turns out it seems to be a disease of blood clotting, that affects the lungs. It makes our treatment that we are doing the wrong thing and potentially even harmful, so we’ve changed a lot
about what we are doing… [Interviewee 1704, ICU Consultant]

It’s a new disease process. We are learning all the time; best practice of evidence is constantly changing so keeping on top of that is difficult. […] Accessing best practice can be difficult, we have a new intranet […] where we have our single page checklist with guidance for these patients that is shared with [redacted]. I was thinking about what sort of things would be helpful and actually because it is a new disease process there is quite a lot of new things we don’t normally do, so a lot of regular blood tests that happen for example at days 1, days 3, 5, 7, keeping on top of when those are. They are all on our daily checklist, on our management guide. They are tests we wouldn’t normally do, but they are looking for specific things. Other new disease processes on top of the COVID, like HLH [hemophagocytic lymphohistiocytosis] that can happen, so sort of screening for those. [Interviewee 1587; ICU Consultant]

Interviewee 1704 (ICU Consultant) further noted the continuous operational changes occurring in the COVID-19 “pods,” defined by Interviewee 1587 (ICU Consultant) as the “designated COVID-19 areas” where additional personal protective equipment (PPE) is required for entry, which created additional psychological and physical strain:

The COVID pod has changed quite a bit because advice on personal protection has changed. Initially, all the patients were in closed rooms with the doors shut and you had to put all the equipment on to go in and see them. […] Now, everybody is wearing it all the time in the hall throughout the unit and the doors are open. So, when the doors were shut you didn’t do the ward round at the bedside, you were stood outside looking at all the data and the nurse inside was writing stuff on a white board to show somebody on the outside to write on the piece of paper. That wasn’t a sustainable solution because it was quite staff intensive.

[when health care staff move into the] high-risk areas […] there is a barrier psychologically to going into the room.

The accounts of participants illustrate the importance for an ICU dashboard to have the capacity to be flexible and easily updated due to constantly changing protocols, information, and advice (eg, reminders to check for specific parameters, changes in protocols regarding closed rooms or designated COVID-19 pods). This could draw from national- and hospital-level advice or information that is pushed out to staff via the dashboards pulling from various NHS information systems. This is a key requirement to ensure the dashboard continues to be usable as we learn more about COVID-19 and to potentially adapt this device into a tool that can remain integrated into ICUs more generally. These narratives highlight the ICU staff needs for an adaptable dashboard that can be updated with constantly changing real-time data about patient parameters, with new and revised routine alerts for new tests, and with reminders for specific trends to look out for when dealing with COVID-19 patients.

A Mobile Dashboard

Interviewees 1252 (Anesthetic Registrar), 1587 (ICU Consultant), 1704 (ICU Consultant), and 1839 (Matron) stated a preference for a dashboard that they would be able to use while walking around the ICU to attend to patients:

There is a lot of walking around by the nurse in charge, just to touch base with people for support and things. They are not static, so it would have to be a mobile solution [Interviewee 1704, ICU Consultant]

A lot of my clinical duties are mobile so to not have the technology follow me and having to use fixed desktops is sometimes quite frustrating. As I say I am mobile, the whole of clinical care is conducted in a very mobile fashion [Interviewee 1252, Anesthetic Registrar]

Thus, the dashboard would potentially replace paper forms and allow for simple and efficient data entry (both qualitative and quantitative) as staff move around while on shift. Interviewee 1704 (ICU Consultant) discussed new issues that arose specifically from dealing with COVID-19, such as the heavy reliance on paper forms in their hospital, given that computers are situated in high-risk areas, which, due to PPE use regulations, makes access to computers difficult. Hence, having a dashboard with remote access to various hospital systems and records that can also be used in mobile devices and taken by staff members outside of the pods would be helpful.

The other thing that COVID-19 has bought in with it, is our clerical staff […] who input a lot of data onto the computer, they’re not able to go into the area because they’re not fit tested with the masks. We are taking pictures of the observation chart with the iPads and uploading them to [redacted] […] for them to then look at remotely, which is a bit of a fudge. It’s not brilliant, but it’s better than nothing at the moment.

By reducing paper forms and collating information neatly in one place, this would ensure information can be carried from patient to patient around the ward and would simplify a series of traditionally offline protocols, task, and data management. While simple in concept, realistically bringing together several diverse information systems and data across hospitals into a unified system or database is a highly complex task and may be difficult to integrate fully into the workforce [41]. However, examples of publicly available, deidentified EHRs do exist, such as the Medical Information Mart for Intensive Care (MIMIC) as part of the Beth Israel Deaconess Medical Center [42].

Customizability and Usability

A key theme arising from the interviews was the users’ needs to customize the dashboard (see Multimedia Appendix 1). For example, Interviewee 1099 (ICU Consultant) expressed a need for customization to support the different tasks and roles of the ICU staff, since clinical information such as patient parameters...
is crucial to performing the Intensive Care Consultants’ tasks, whereas operational data such as duration of patients in prone position would not necessarily be of interest to doctors yet would be of great importance for nurses. Similarly, Interviewee 1252 (Anesthetic Registrar) suggested the usefulness of knowing the patients’ pending and past procedures. Hence, the data required for a dashboard to be useful are extensive, complex, and would draw from several hospital information systems, including (where * denotes parameters that were also mentioned as important to view over time as trends and ^ denotes markers particularly important and of interest for COVID-19 patients): C-reactive protein (inflammation marker)^*, D-dimers (blood clotting marker)^*, ferritin (inflammation marker)^*, lymphocyte count (inflammation)^*, platelet count (inflammation)^*, procalcitonin levels (inflammation)^*, pending tests for specific HLH patients, blood pressure^*, white blood cell count^*, fraction of inspired oxygen, oxygen level^*, oxygen supply level for personalized care, peak airway pressure for COVID-19, tidal volume size, positive end-expiratory pressure^*, arterial oxygen/inspired oxygen ratio (PF ratio), plateau pressure^*, type of ventilation, Sequential Organ Failure Assessment (SOFA) score, Glasgow Coma Scale score, intracranial pressure^*, ventricular tachycardia^*, infusion rate of vasopressors^*, number and absorption of nutrition calories, and COVID-19 status. Further, specific requests for alerts to be associated with specific data were noted, such as abnormal values across parameters, 7-10 days of static oxygen (meaning a computed tomography pulmonary angiogram scan can be completed), pending tests for COVID-19 patients, clinical deterioration, and pending procedures for patients. Finally, other information was requested that was more operational in nature, including bed layouts, number of patients on dialysis, nurse locations, patient numbers across units, patient flow information (eg, admission, discharge, changing of units), and relevant patient handover information.

Unsurprisingly, a lot of data and information were requested to appear in the dashboard, which may be overwhelming and difficult to navigate or simply not of relevance or interest to certain roles in the ICU. Interviewee 1099 (ICU Consultant) suggested that a fully individualized dashboard for each staff member would be ideal as this allows for a tailored configuration for each of the roles within the ICU. However, it is important to be cautious when implementing highly customizable systems, since this customizability may induce errors when overstressed staff are required to make fast decisions. Hence, careful consideration is needed to determine what degree of customization is advisable for this device at the user level. It is important to have functionality that allows additional staff access on an ad hoc basis. As Interviewee 1587 (ICU Consultant) stated, staff from different areas of the hospital may need access to the dashboard. For example, COVID-19 patients often need nutritional and dietary assistance; hence, nutritionists may also need specific systems access for these patients.

Alternatively, Interviewee 1839 (Matron) suggested having a split view where parameters could be presented as broader categories such as “Clinical Parameters” (eg, ventilation, tidal volumes, dialysis) or “Safety Parameters” (eg, delirium, infection, prone position turns), so that these measures could be useful for various health care staff functions and responsibilities. However, this could also be problematic if functionality allows for toggling role-specific parameters and the lack of prima facie data may cause staff to miss trends in parameters not shown on the screen.

Several participants noted the importance of integrated systems (see Multimedia Appendix 1). For example, Interviewee 1099 (ICU Consultant) stated their current systems necessitates that they “have to open up 5 different screens to get the data [they] need and that is pretty labor-intensive,” or that they are “often chopping and changing through different programs…” which demonstrates the importance of having an intuitive interface that consolidates relevant data on demand. Similarly, in terms of usability, it is problematic if a system requires additional “administration” for staff to find information and results on pending tests. This can be illustrated with the following account from Interviewee 1099 (ICU Consultant) about the [redacted] system:

When you have somebody who comes in with a chest infection, 15 tests are ordered: five aren’t back, five were never sent, and five are back, but you would never know that by looking at [a system]. You would only know that by looking at this separate program.

It is therefore crucial for the dashboard to be well-integrated with other hospital systems to avoid data and work duplication. It is thus paramount to ensure that a new dashboard does not add complexity but reduces workload to access data by extracting it from the existing systems. It is essential for staff to be able to customize their view to quickly sift through large amounts of information, understand patient needs, and determine next steps. As a requirement, this is achievable and realistic, as it is common for information systems to have user profiles with individual logins [43]. Further, having default user profiles based on roles and grades within the workforce is a reasonable requirement to implement, where individuals can request additional accesses ad hoc. However, the level of customizability offered alongside user profiles would require additional testing and research (eg, changing color schemes, data access changes).

**Dashboard Layout and Trends of Incoming Data**

**Patient Overview**

In general, there was a tendency for all participants to comment on how data should be processed and presented on the dashboard. Interviewee 1839 (Matron) stated a preference for the ability to view the whole ward (which could also track bed and patient expansion). If this were to match the physical layout of the beds, this would be useful to find patients quickly and to effectively plan a patient’s acuity (level of nursing care) quickly. This is especially important should the unit become busy or indeed require a quick and large-scale expansion in patient numbers, as envisaged at the UK’s “Nightingale” hospitals. In addition to the ward view, Interviewees 1099 (ICU Consultant), 1704 (ICU Consultant), 1839 (Matron), and 1252 (Anesthetic Registrar) suggested the inclusion of a summarized “Condition at a Glance” view, which would allow for those starting their shift to quickly get up to date. This was echoed by Interviewees 1704 (ICU Consultant) and 1839 (Matron), who suggested a display of “overarching” parameters of all patients on the ward.
Interviewee 1587 (ICU Consultant) reported a similar requirement that would display the most critical parameters (e.g., SOFA scores, tidal volume), which would allow the Intensive Care Consultants to see the trend of a patient’s current condition. Interviewee 1587 (ICU Consultant) argued that calculating SOFA scores is an arduous task for junior doctors; hence, making these calculations available and easy to interpret in a dashboard will save ICU staff time and cognitive energy [44].

Of course, there is more to ICU patients’ health care than a first-glance interpretation of these parameters. Interviewee 1839 (Matron) stated the complexity of patient monitoring when patients might appear fine in terms of typical baseline metrics, but in reality, their actual state is misrepresented by data: “[a] patient could be fine, but they’re on a lot of inotropes (eg, noradrenaline) or ventilation, and it looks OK, but they are on 100% oxygen and/or quadruple noradrenaline.” The intuitive assessment required in such cases could only be achieved through an appropriate identification of the daily and hourly trends by the ICU staff, of the patients’ inflammatory markers and oxygen levels, according to Interviewees 1704 (ICU Consultant) and 1587 (ICU Consultant). Thus, the availability of these on a dashboard will be particularly important for health care staff to decide on COVID-19 patients’ clinical care and to devise provisions for their safety, as expressed by Interviewees 1099 (ICU Consultant) and 1704 (ICU Consultant). Further, Interviewee 1587 (ICU Consultant) discussed the significance of data trends, especially regarding COVID-19 patients, such as decisions about when patients can be weaned off ventilation as well as monitoring the prone position ventilation timings. Additionally, a dashboard could track specific sets of data, which may offer statistical insights (Interviewee 1252, Anesthetic Registrar) into how to better treat future patients for particular diseases or conditions, as these data will be able to provide baselines and expectations, particularly for new diseases such as COVID-19.

Data Visualization, Warnings, and Alerts

There were differences among ICU staff as to how they wanted data to be presented in this type of tool (see Multimedia Appendix 1). A suitable example of this would be the informants’ preferences for graphical displays that would help to address two important issues pointed out by Interviewee 1587 (ICU Consultant):

> I think if you click on it, you can see a graph, but to be honest we don’t regularly do that. Ninety percent of the time it’s just numbers completed on a sheet…We literally have a piece of paper that junior doctors fill in in the morning or if they haven’t, we go on [redacted] and click on the CRP trend on there and see what’s happening. You can get a graph of it, but it’s not ideal.

Hence, it is important to enable users to have a degree of autonomy to customize the dashboard for individual patients (Interviewee 1099, ICU Consultant), and to transform data and information into a format that best suits their learning needs and information processing style. Alongside ways to visualize data, it is critical to have suitable real-time deterioration alerts both for clinical and safety parameters, which may include visibly highlighted alerts on abnormal values, as well as real-time alerts on staff deviations from practice (Interviewees 1099 [ICU Consultant] and 1704 [ICU Consultant]). However, when it comes to the display of these warnings, Interviewee 1099 (ICU Consultant) discussed what would be the most suitable parameters for COVID-19 patients (e.g., PF ratio and driving pressure) to help clinicians decide when to start weaning the ventilators, as well as the adequate time parameters for the graphic view. This interviewee stated the importance of including suitable “cutoffs” in terms of data presentation, because to plan COVID-19 patient care, Intensive Care Consultants must consider both the presence of abnormal values as well as how these values behave over time (trends). In addition, Interviewee 1839 (Matron) noted determining thresholds for colors and notifications is not a trivial matter, especially if they need to be tailored to specific medical conditions or personalized to each individual patient. Therefore, a dashboard that presents data in a way so that the user could see a longer patient history may be extremely helpful. This could be facilitated by allowing the graphs to be scrolled through horizontally to show earlier data and see longer longitudinal trends since the patient’s admission.

It is important to be aware that notifications or alarms are extremely common in ICUs due to the variety of abnormal values of health parameters in critically ill patients. Therefore, when designing a new dashboard, a reasonable balance is needed to avoid “alarm fatigue” and to prevent staff from missing patient deterioration markers, which can lead to detrimental outcomes [45-48].

The dashboard would give me the triggers to go sniffing around into the detail of the data […]. I would just highlight the noradrenaline box and the base excess [and hide other parameters and] look at those two things. What’s the trend? […] I think the personalization of being able to manipulate it on one screen… [Interviewee 1099: ICU Consultant] trends are brilliant. […] it doesn’t really matter what the noradrenaline is, if it has doubled in the last hour, it’s not a good thing. […] Actually, it’s the step change that is the important thing. For me, I was straight drawn into the color change and the arrows, then just lost sight of the numbers a bit. That is probably a good thing because I would then go looking into that patient detail on the system to see why their noradrenaline is going up and doubled. So that was quite a good trigger: [Interviewee 1099; ICU Consultant]

Here, the requirement relates to the layout of the dashboard and how data are presented (with some in real time). This called for flexibility in terms of how data can be graphically displayed to suit the staff member using the device, which may also help with the number of alarms and notifications in the ICU. This requirement is relatively simple to implement within a dashboard system, where users will be able to shift between graphic and tabular displays of information or seeing longer-term trends of a patient, for example. However, attempting to visualize data and highlight when parameters are shifting negatively for the patient is inherently more complex; thus, testing and examining
what the thresholds should be are crucial to reduce alarm fatigue among additional stress for staff. This would include investigations regarding individualistic measures versus overall baseline “cutoffs.”

**Task and Staff Management**

**Patient Handovers**

A key theme regarding staff and task management were patient handovers (see Multimedia Appendix 1). Handovers (both regarding staff shifts and turnover of their allocated patients), are complex two-way processes between a variety of staff signing out and updating those coming in to take over, where the accuracy and effectiveness of this information exchange “will facilitate consistency and continuity of care” [49]. This is particularly important for critical care patients, where omission of pivotal information during the handovers could influence future treatment and subsequently cause failures in patient management [49]. Since the use of a dashboard could rapidly help to capture and track wider information regarding patient status and care requirements, it is evident that the implementation of this tool in the ICU environment could facilitate more structured and effective patient handovers. For example, Interviewee 1839 (Matron) stated:

> We know that handover time and transfer of care is a pinch point where if there is going to be an issue or problem occurring, we often track it back to that point in time. Where something has been missed, not handed over or at that point they may look at something and go “that’s not what I remember it being.” That’s the trigger to go back systematically through all their different bits. Or a doctor has come along and changed the rate of a pump and not told somebody. We know that is a really pivotal time so some kind of overarching view of the main clinical elements of a patient care would be helpful. That would give them a visual aid to that and anything that would help a hand over of care, would absolutely be welcome.

Similarly, Interviewees 1099 (ICU Consultant) and 1839 (Matron) coincided in stating the importance of allowing time for scrutiny of new inputs of their colleagues and data updates from various patients that have been handed over. However, as Interviewee 1099 (ICU Consultant) pointed out, outlining changes in patient parameters during quick handovers is done with great difficulty while having to navigate multiple hospital systems to gather the information required:

> I had a handover from my colleague - but I want to process it in my own mind and want to see what’s changed over the last 12 hours, since they handed over; [currently] I have to open up 5 different screens to get the data I need and that is pretty labor intensive.

**Data Entry**

Due to the important consequences at stake, extreme care and monitoring are exercised in the ICU environment to ensure accurate data input in their systems.

If you are feeling responsible for the patient, which you are as a consultant, you need to double check that [data, patient notes]. The only way to do that is to physically look at those [systems] and paper notes yourself [Interviewee 1099, ICU Consultant]

Moreover, when planning patient care targets, ICU staff in managerial positions must carefully balance patient-management workloads of staff with their data input tasks. This links in with comments from Interviewee 1099 (ICU Consultant) when expressing concern about overtired staff with data entry responsibilities such as recording general observations about patient progress and invasive procedures, among others:

> [staff] will be absolutely knackered at three in the morning and just put [in] the bare minimum. They’ll go, “patient had an operation, and this is what happened” and forget about other staff.

On these grounds, participants considered having a tool that can assist staff in the transcription and modification of patient data with a minimum error rate to be important. Arguably, errors may continue to occur in the presence of a digital interactive dashboard; however, research has shown that using digital systems to collect and log data (rather than pen and paper) reduces errors in data recording and data entry [50,51]. The use of a dashboard helps to provide a faster way to populate handover or debrief notes. Furthermore, Interviewee 1252 (Anesthetic Registrar) noted the lack of any formal system at present to document any medical advice provided to patients over the phone. This could be a simple note about the patient that can be added to the dashboard to ensure an overview of all advice and information previously provided to the patient.

**Task Management**

A major issue reported by the participants was a lack of warnings in their current systems about forthcoming completion times of pending tasks and targets, which could be built into a dashboard. Interviewee 1099 (ICU Consultant) stated that a careful balance must be struck with off-target warnings to avoid undermining staff confidence:

> The warnings that are built into the target need to be in advance, there is no point telling people at midnight you’ve not met your fluid balance target because it will just demoralize people.

Consequently, it is unsurprising that one of the participants’ most frequently mentioned dashboard requirements was to have warning notifications ahead of completion times (see Multimedia Appendices 1 and 2), which would certainly work as a task management system that will ensure timely completion of the multiple pending tasks and daily targets of medical and nursing ICU staff, including ventilation weaning, daily prone and supine ventilation sessions for COVID-19 patients, and other safety tasks of nursing staff; invasive procedures such as tracheostomies; monitoring pending microbiology and specific tests for COVID-19 patients; changing of drugs; and speaking to relatives (Interviewees 1099 [ICU Consultant], 1159 [Sister], 1252 [Anesthetic Registrar], 1587 [ICU Consultant], 1704 [ICU Consultant]), among many other responsibilities. Interviewee 1099 (ICU Consultant) stated that this type of task management system would give managerial staff peace of mind by knowing
that “loops are closed,” especially when the ICU becomes extremely busy and “people forget about minutiae.” Furthermore, in a context of ever-changing guidance and information, Interviewees 1159 (Sister) and 1587 (ICU Consultant) also agreed that it would be extremely useful to have daily task checklists (eg, safety checks) as a suitable requirement of the dashboard, which should also include enabling inputs of data as needed. Interviewee 1587 (ICU Consultant) also pointed out how having a dashboard to prompt staff members to finish tasks would be important to address the absence of warnings on pending targets in their current system, especially with new diseases such as COVID-19 when more tests than usual are frequently needed. For example, Interviewee 1099’s account illustrates how ICU staff struggle to juggle their immediate tasks with their daily patient care targets for both COVID-19 and non-COVID-19 patients:

*There will be a patient that will be on multiorgan support, have to go for a scan, have to go to theatre, come back. Then the nurses will try to make sure the patients are fed, […]. all sorts of complex care issues [are] going on. That [eg, fluid] target then drifts into the background. There’s no prompt to say, “your fluid balance is nowhere near target and you have four hours to go. What are we going to do to solve this problem?”*

To obtain a general picture of the ICU patient care flow, interviewees discussed the advantages of having an overall view of staff numbers and their corresponding workload, alongside data about patient admissions and patient flow (eg, patient discharge and transfers either for tests or to other wards), as pointed out by Interviewees 1159 (Sister), 1839 (Matron), 1704 (ICU Consultant), and 1252 (Anesthetic Registrar). For example, the following testimony of Interviewee 1159 clearly portrays the managerial ICU staff need of a well-updated and integrated system showing workload allocation:

*We need a way of knowing which nurse is in each bed space so if there is an issue, we can speak to that nurse looking after that patient. If it was two shifts down the line, there was something we needed to get hold of somebody about.*

At the same time, from the following account of Interviewee 1704 (ICU Consultant), it can be inferred that managerial staff are simultaneously responsible for overseeing the staffing of the unit while monitoring the changing conditions of all patients. Hence, the requirement for a new system to collect these two categories of data was mentioned:

*They collect this data on a piece of paper as well and it’s in pencil so they can rub stuff out and change it. I have been trying for years to get iPads for them to use, we need some kind of software for that. That data would be really important to analyze: the patterns of activities during the day, to optimize our staffing models or things like that.*

This statement highlights that the implementation of a dashboard for ICU use could provide a much-needed opportunity to shift from pen and paper to a digital system of data collection and monitoring, alongside a new strand of data analysis that could help optimize staff time and workload. Interviewee 1252 (Anesthetic Registrar) provided additional corroboration of a dashboard’s value to mobile staff such as Registrars by having an effective oversight of all patients pending transfers to other wards and for tests (eg, for a computed tomography scan or other tests). This relates to Interviewee 1704’s (ICU Consultant) statement about the need to better optimize staffing models and the daily distribution of tasks by accurately monitoring staff workload and whereabouts as follows: tracking staff timings for patient care for each allocated patient, as well as producing continuous insight into the location of the health care staff throughout the different ICU wards during their shifts (Interviewees 1159 [Sister] and 1252 [Anesthetic Registrar]).

According to Interviewee 1252 (Anesthetic Registrar), this two-fold patient-staff tracking system would be very useful for staff who are constantly busy (Registrars and Running Nurses). With this new system, they can efficiently share their work in relation to their location, the patient transfer destinations, and the numbers of daily transfers.

This final requirement, which touches on all prior requirements for the dashboard, is ensuring that the dashboard is seamlessly integrated with other hospital systems. In this way, staff can access additional, external data (eg, authorized views of blood test results from other hospital systems). This would help to tackle the issue described by Interviewee 1252 (Anesthetic Registrar) when characterizing the process of patient admissions from other hospitals as “data heavy” with a lot of “transcribing of various different sources onto the intensive care unit systems,” where data are currently not being pulled neatly into one system. As stated previously, this is a highly complex task that would be difficult to integrate as current systems are siloed across hospitals [42].

Concerns About Dashboard Use

As shown in Multimedia Appendix 2, there were two main operational concerns among ICU staff regarding the use of dashboards. The first concern, raised by Interviewee 1587 (ICU Consultant), related to the potential increase of infection via use of mobile technology and equipment. On the assumption that these dashboards will be used in mobile devices such as iPads, Interviewee 1587 (ICU Consultant) questioned whether these devices should be allowed into the infection-controlled areas of the ICU, specifically at the bedside of COVID-19 patients. As possible solutions, Interviewee 1587 (ICU Consultant) mentioned the use of disinfecting wipes for mobile devices, but mostly adopting clinical protocols that avoid the need for using a device at the bedside.

*We don’t take the iPads in to see patients. In terms of risks, they are potential fomites, a sort of vector for transmission of infection. We wouldn’t take them into bed spaces. Much like cleaning our phones, we are good at cleaning with special Clinell wipes. Yes, there is a potential risk… We stand outside where it is lower risk. We cluster round as a ward round to write up our notes and decide the plan. Only one of us will go in to examine the patient. Everyone else will wait outside.*
The second operational concern regarding dashboard use related to the differing levels of technology literacy among ICU staff. For example, Interviewee 1099 (ICU Consultant) stated that unless there already is a culture around the use of dashboards and technology, encouraging staff to actually use and engage with this type of tool might be difficult [27]. This is an interesting comment as Interviewee 1099 is based at the more technology-enhanced ICU that uses dashboards among other devices regularly, where such tools are integrated. Hence, these concerns additionally relate to the managing expectations in staff (Interviewee 1587, ICU Consultant) of what the dashboard will do, how it should be used, and protocols regarding these devices, as it is important not to oversell a new technology’s potential impact for the ICU.

Finally, there was some ambivalence from staff members about the dashboard design, as illustrated by the concerns raised by Interviewee 1099 (ICU Consultant), who is based at an already technology-enhanced ICU, in relation to what should be the “acceptable” parameters for the notification and warning timings with regard to both COVID-19 patients and staff targets. Interviewee 1099 (ICU Consultant) also raised concerns regarding notifications or alerts coming in at inappropriate times, since this could cause the adverse effect of “demoralizing people.” This participant further mentioned a concern for having a tool with which staff could compare their unit target achievements with other units. This could increase ICU staff stress, and might lead to suffering from burnout or cognitive overload, “wow, mine’s [targets or parameters] all red, things aren’t going very well,” which could have ramifications for both patients and staff due to the well-documented fact that the ICUs are incredibly stressful environments [5,6,52]. These are important concerns to address early and to ensure staff are all fully informed regarding the system itself and the transparency regarding how the digital logs that it will produce may be used outside of patient monitoring (eg, can these be used to assess staff performance in the workplace?). Hence, when implementing new systems, the engagement of end users is key to ensure expectations are set and staff can feel supported by these new systems.

Discussion

Principal Findings and Conclusions

In response to the critical situation of two local ICUs, we conducted a series of interviews to elicit requirements for a bespoke dashboard to help ICU staff save time and work more efficiently, particularly during the COVID-19 pandemic. We found that despite having limited access to end users, our approach of conducting remote requirements interviews for developing a dashboard for COVID-19 ICUs has been successful. The rapid cycle of interviewing end users, prototyping the user interface, and iterating over the software design, despite taking place in extreme and distressing circumstances of the pandemic, has proven to be an effective way of producing functional software requirements. These requirements have in turn allowed for the development and deployment of an interactive dashboard currently being tested and evaluated across two hospitals.

The first requirement was the need for a flexible dashboard, primarily to help ICU staff respond to rapidly changing guidance for the management of this new disease. The second requirement emphasized the need for a mobile dashboard, which allows staff to walk around wards with real-time data and information of patients. The third requirement focused on customizability of a dashboard, stemming from the great diversity of roles and tasks conducted by ICU staff. Related to this was the fourth requirement, which was the ability to track and visualize real-time data and daily/hourly trends on patient parameters. The fifth requirement was aimed at pending tasks and targets for staff management. All requirements highlight a need for the integration of different hospital systems within the dashboard, which is a longstanding challenge in medicine [41]. Alongside these requirements, participants raised concerns regarding the infection-risk safety issue of bringing devices into the ICU and of the timing of warnings and alerts.

The study findings confirmed that digital solutions for ICU use would potentially reduce the cognitive load of ICU staff and reduce clinical errors at a time of notably high demand of intensive, critical health care [17]. As summarized by Interviewee 1099 (ICU Consultant), the beneficial implications of having this dashboard would hopefully be that “not only will it make the system more efficient” but it will further give them the possibility of “looking after more patients more safely.”

Limitations

We acknowledge that the sample size is small due to the workload on ICU staff caused by the COVID-19 pandemic, which was the underlying motivation for this research. However, we did capture requirements, perspectives, and experiences from a wide range of clinical roles within the ICU environment across two somewhat different hospitals, particularly from those heavily involved in the health care of COVID-19 patients during the pandemic. There are a number of ways we could have elicited the requirements for the dashboard, including from questionnaires, joint application development, storyboarding, and protocol analysis [31,32]. However, with the time pressure to develop a dashboard that was working and usable, alongside the time pressures ICU staff were under, we wanted to continue with the most straightforward and least cognitively heavy method of elicitation for the ICU staff.

Further, we are aware that our dashboard will require extensive testing in the ICUs with end users such as our interviewees to refine the design and functionality. This would include examining how we would tailor the dashboard for different roles (eg, having a home page with various types of information for each role such as nursing staff vs a consultant or a nutritionist). This will be an iterative process, where we acknowledge that not all needs will necessarily be met; however, our aim is to ensure the device is usable and enhances staff.

We are also aware that our sample comes exclusively from the region of Bristol, which may not be representative for ICUs across the rest of the United Kingdom or indeed outside of the United Kingdom. However COVID-19 has impacted health care provision in many regions and many countries worldwide, and many of the staff, patient, and task management requirements; the ability to track and monitor trends; and the
Dashboard customization for individual staff members are likely to be common requirements both across the United Kingdom and around the world [53].

Acknowledgments
This work is partly supported by the Engineering and Physical Sciences Research Council Digital Health and Care Centre for Doctoral Training at the University of Bristol (UKRI grant EP/S023704/1 to MW). CMW is funded by the South West Better Care Partnership, supported by Health Data Research UK.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Code tree classification of dashboard information preferences (96 codes).
[XLSX File (Microsoft Excel File), 12 KB - humanfactors_v9i2e30523_app1.xlsx ]

Multimedia Appendix 2
Code tree classification of concerns about dashboard use (24 codes).
[XLSX File (Microsoft Excel File), 10 KB - humanfactors_v9i2e30523_app2.xlsx ]

References


**Abbreviations**

- **EHR**: electronic health record
- **HLH**: hemophagocytic lymphohistiocytosis
- **ICU**: intensive care unit
- **MIMIC**: Medical Information Mart for Intensive Care
- **NHS**: National Health Service
- **PF**: arterial oxygen to inspired oxygen ratio
- **PPE**: personal protective equipment
- **SOFA**: Sequential Organ Failure Assessment

---

**Edited by A Kushniruk; submitted 19.05.21; peer-reviewed by A Frahling, T Aslanidis, M Wright; comments to author 08.07.21; revised version received 27.09.21; accepted 03.01.22; published 13.04.22.**

Please cite as:

Davidson B, Ferrer Portillo KM, Wac M, McWilliams C, Bourdeaux C, Craddock I

Requirements for a Bespoke Intensive Care Unit Dashboard in Response to the COVID-19 Pandemic: Semistructured Interview Study

JMIR Hum Factors 2022;9(2):e30523

URL: https://humanfactors.jmir.org/2022/2/e30523

doi: 10.2196/30523

PMID: 35038301
A Participatory Design Approach to Develop Visualization of Wearable Actigraphy Data for Health Care Professionals: Case Study in Qatar

Kamran Khowaja1,2, PhD; Wafa Waheeda Syed3, MSc; Meghna Singh4, MS; Shahrad Taheri5,6,7, PhD; Odette Chagoury5,6,7, PhD; Dena Al-Thani1, PhD; Michaël Aupetit3, PhD

1 Information and Computing Technology Division, College of Science and Engineering, Hamad Bin Khalifa University, Education City, Qatar
2 Department of Computer Science, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Hyderabad, Pakistan
3 Social Computing, Qatar Computing Research Institute, Hamad Bin Khalifa University, Education City, Qatar
4 Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, United States
5 Department of Medicine, Weill Cornell Medicine, Doha, Qatar
6 Department of Medicine, Weill Cornell Medicine, New York, NY, United States
7 National Obesity Treatment Center, Qatar Metabolic Institute, Doha, Qatar

Corresponding Author:
Michaël Aupetit, PhD
Social Computing, Qatar Computing Research Institute
Hamad Bin Khalifa University
Research Complex Building 1
Education City, 34110
Qatar
Phone: 974 44547150
Email: maupetit@hbku.edu.qa

Abstract

Background: Several tools have been developed for health care professionals to monitor the physical activity of their patients, but most of these tools have been considering only the needs of users in North American and European countries and applicable for only specific analytic tasks. To our knowledge, no research study has utilized the participatory design (PD) approach in the Middle East region to develop such tools, involving all the stakeholders in the product development phases, and no clear use cases have been derived from such studies that could serve future development in the field.

Objective: This study aims to develop an interactive visualization tool (ActiVis) to support local health care professionals in monitoring the physical activity of their patients measured through wearable sensors, with the overall objective of improving the health of the Qatari population.

Methods: We used PD and user-centered design methodologies to develop ActiVis, including persona development, brainwriting, and heuristic walkthrough as part of user evaluation workshops; and use cases, heuristic walkthrough, interface walkthrough, and survey as part of expert evaluation sessions.

Results: We derived and validated 6 data analysis use cases targeted at specific health care professionals from a collaborative design workshop and an expert user study. These use cases led to improving the design of the ActiVis tool to support the monitoring of patients’ physical activity by nurses and family doctors. The ActiVis research prototype (RP) compared favorably with the Fitbit Dashboard, showing the importance of design tools specific to end users’ needs rather than relying on repurposing existing tools designed for other types of users. The use cases we derived happen to be culturally agnostic, despite our assumption that the local Muslim and Arabic culture could impact the design of such visualization tools. At last, taking a step back, we reflect on running collaborative design sessions in a multicultural environment and oil-based economy.

Conclusions: Beyond the development of the ActiVis tool, this study can serve other visualization and human–computer interaction designers in the region to prepare their design projects and encourage health care professionals to engage with designers and engineers to improve the tools they use for supporting their daily routine. The development of the ActiVis tool for nurses, and other visualization tools specific to family doctors and clinician researchers, is still ongoing and we plan to integrate them into an operational platform for health care professionals in Qatar in the near future.

https://humanfactors.jmir.org/2022/2/e25880

JMR Hum Factors 2022 | vol. 9 | iss. 2 | e25880 | p.104
(page number not for citation purposes)
participatory design; user-centered design; visualization; health care professional; persona; brainwriting; heuristic walkthrough; use case; interface walkthrough

Introduction

According to the World Health Organization (WHO) report of 2018 [1,2], lack of physical activity is the fourth leading risk factor for mortality. Physical activity reduces the risk of coronary heart disease, stroke, hypertension, depression, type 2 diabetes, and several types of cancer. Unfortunately, physical activity across many countries is declining. In the context of Qatar, researchers at Weill Cornell Medicine - Qatar (WCM-Q) conducted a study among elementary school children between ages 7 and 12 [3]. The authors found that 42.1% of these children were either obese or overweight, and their sleep was significantly shorter than children with a healthy weight. In another study on prevalent health issues among Qatari citizens and long-term residents [4], the authors found that 83% of the population undertook little to no physical activity, and almost half of the population did not do any physical exercise. Hence, there is a need to increase the physical activity of the Qatar population to reduce the risk of related diseases as mentioned in the WHO 2018 report.

Many behavioral modification programs have been developed for more than 2 decades to reduce physical inactivity [5-8]. Nowadays, technologies allow continuous recording of individual physical activity over several days. Moreover, the use of smartphones and wearable devices (smartwatches, wristbands, etc) among children, adults, or older adults has increased in the last decade. Smartphones and wearable devices are then actively used to record, measure, and monitor body movement and activities performed by an individual using a global positioning system and accelerometer installed on these devices [1,9-11]. The visualization of the recorded activity data can then show the time when an individual was the most or least active throughout the day [12] and support monitoring and exploration of such activities. We focus on the design of such visualization tools in this work.

There is a growing trend in visualization studies to explore ways to represent wearable data for self-monitoring sleep [13], analyze data by health coaches [14] or researchers [15], evaluate performance dashboard for sport [16], or evaluate time-based activity graphical representations on mobile phones [17]. Other studies explored the best approach to visualizing the data to support behavior change [18,19] or provided a visualization dashboard to help patients understand their longitudinal health data [20]. Still, the visualization of wearable data is an active research area. A natural approach to start with is to repurpose existing visualization tools such as the Fitbit Dashboard to visualize data in a healthcare setting. However, the actual needs of health care professionals may depart substantially from the ones of the general public self-tracking their physical activities. To the best of our knowledge, there is no visualization tool specifically designed to support the health care professionals in monitoring and analyzing physical activities of patients through their wearable actigraph data. We also could not find a set of use cases and user roles covering such needs.

Moreover, several studies [21-23] have demonstrated the importance of the cultural, social, and local context when designing medical or health care technological solutions. Despite this view, the literature on technology acceptance mostly concentrates on highly developed North American and European countries, and little is known about health technology use and data visualization in the Arab world, including the Gulf countries [24-29] such as Qatar [3,4]. Arab countries share lots of similarities, such as cultural and religious values, language, and lifestyle [30,31], and are quite different from North American and European countries. Salgado et al [32] has highlighted that culture plays a vital role from the investigation to the design or development of new methods, theories, techniques, and systems. Hence, cultural specificities were expected when we started this project and we decided to follow a participatory design (PD) approach to collect the potentially culturally specific needs of end users.

Alabdulqader et al [33] highlighted a need to reduce the cultural gap between technology designers and users by using a PD approach. PD aims to design solutions that consider the local context and culture and has been used effectively in the health/medical domain [34-44]. PD allows researchers to involve potential users of a product or technological solution in the ideation, design, development, or appropriation of the solution [35]. Kanstrup et al [35], as a part of their review, found that workshop/group sessions/focus groups, interviews, and prototyping have been more commonly used in PD sessions of health information technology. We followed this approach in our studies. The use of opportunistic research and sampling is commonly used in health care research as it allows researchers to use the available participants or research instruments to perform research chores [45-49]. To the best of our knowledge, there is no interactive tool that has considered the needs of local health care professionals in Qatar in their regular activities. These activities include understanding and monitoring their patients and helping/assisting them to improve their physical activity, sleep, and eventually reduce obesity. Results from a previous study [4], informal discussion with the authors [3], and an approach of opportunistic research were used as a basis to design the first prototype of an interactive tool (ActiVis) to support the mentioned needs of the local health care professionals.

This paper reports on the PD and summative evaluation of a second version of the ActiVis prototype to visualize activity data from wearable devices, which meets the needs of local health care professionals for monitoring the physical activity of their patients, to improve the physical activity of the Qatar population. We use methodologies from user-centered design through their wearable actigraph data. We also could not find a set of use cases and user roles covering such needs.
and PD for the first time to design eHealth data visualization in Qatar.

**Methods**

**Research Protocol and User Studies**

**Overview**

The users and their needs increased over time as studies were conducted as a part of this research and ActiVis was accordingly modified and reported in different sections of this research. The development of any technological solution is not an easy task. It requires gathering and analysis of considerable data from the ideation to the design, development, evaluation, and deployment of the technology. It becomes even more challenging when the local context needs to be considered and incorporated into the technology. The data collection and analysis methods vary from one study to another due to various constraints such as the availability of the target users and the initial uncertainty in the direction of the project, which is refined progressively through the development cycles.

Figure 1 shows the timeline of this work, the studies conducted with their target audience, the methods used, and the venues where they took place. The RPs developed and the user studies (UX) conducted are reported in Textbox 1.

![Timeline of the research work](https://humanfactors.jmir.org/2022/2/e25880/)

**Textbox 1.** Research protocols and user studies conducted.

- RP1: The first research protocol (RP) of ActiVis was developed out of a previous design expert analysis of the requirements not reported here.
- UX1: The first user study (UX) was a workshop conducted with nurses at Hamad Medical Corporation (the largest public health care provider in Qatar) to gather detailed requirements, personas, and usage scenarios, to design and develop the second RP of ActiVis (RP2) together with a set of 6 use cases targeting health care professionals.
- RP2: A total of 3 UX (UX2.1-UX2.3) were conducted to evaluate RP2 on 3 of these use cases. Each study was targeted at 1 type of user as follows:
  - UX2.1: First, an expert evaluation was conducted with clinical researchers at Weill Cornell Medicine - Qatar (WCM-Q). UX2.1 supported improving the descriptions of the use cases, determining which type of health care professional users among nurses, family doctors, and clinician researchers were the actual targets, and evaluating RP2 based on the use cases targeted at nurses and family doctors. Usability issues were also identified as a part of that study.
  - UX2.2: Then, an expert evaluation was conducted with a family doctor visiting Qatar Computing Research Institute (QCRI) to evaluate the second prototype RP2 based on use cases specific to that role as identified from UX2.1.
  - UX2.3: Lastly, a workshop was conducted with nurses from Hamad Medical Corporation. The purpose was to evaluate RP2 on the use case specifically targeted at nurses and to compare RP2 with the Fitbit Dashboard as it provided similar functionalities. The study would allow researchers to understand the differences between both dashboards from the participants’ perspective and improve ActiVis based on their feedback. In this study, Fitbit was used as a comparison because it has a well-thought design [59,60] with similar functionalities required to support the user tasks, and it was the leading wearable technology in the consumer market at the time of the study [61].
- RP3: These studies (UX2.1-UX2.3) led to the design specifications for a third RP not reported here.

The protocol of the studies is described in the remaining subsections, while the results of each study are presented in the “Results” section.

**RP1: Visual Analytic Tool for Actigraphy Sensor Data**

In 2016, one of the authors (MA) started working on a visualization dashboard of wearable data for clinical decision making by health care professionals. This dashboard is aimed at supporting patients to move toward a healthier lifestyle based on their physical activity data. Figure 2 shows parts from the different screens of the initial visualization dashboard (ActiVis) developed as an RP (RP1) based on extensive discussions with health care professionals having expertise in childhood obesity and diabetes in Qatar. The data and initial user needs to be used to design the first prototype were collected as a part of a previous research project [62,63]. The details of RP1 reported in this paper are presented in the “RP1” subsection of the “Results” section.
UX1: Users’ Evaluation Workshop 1 With Nurses

Overview
A first user experiment (UX1) was conducted with the nursing staff of Hamad Medical Corporation on May 2, 2018. The workshop was conducted to gather some of the potential users to generate ideas for the prototype taking the local needs into account. The objective of the workshop was to learn about nurses’ perception of how visual analytics may enable them to promote lifestyle change and provide better advice to patients based on the activity data that would be collected from the patient’s wearable (smartwatch). The session was focused on patients with type 2 diabetes. It included a presentation followed by a brainwriting session, where nurses in groups provided their input on desired information and computer technology solutions to support patient lifestyle changes.

Findings
Our analysis of the data collected from this study led to the design of 6 use cases, and the corresponding user tasks led to the technical specifications of the visualization design that we implemented in the second prototype (RP2) of ActiVis. It is to be noted that use cases were developed from the perspective of nurses who are one of the potential users of the ActiVis tool. However, it was not clear if the description of each use case was adequate or required some improvement, and if all the use cases would need to be implemented in the ActiVis tool, justifying the needs for another set of UX (UX2.1).

Participants
A total of 45 male and female participants, which included nurses as well as nursing informatics professionals working at Hamad Medical Corporation-Qatar (HMC-Q), attended the workshop.

Study Protocol
The nursing staff working at HMC-Q were recruited through an announcement by the chairperson of the nursing department, inviting them into the workshop as shown in Figure 3 to contribute to the development of the health care solution. The participants were split into 4 groups (10-12 members in each) for the brainwriting activities. Each group was provided with a flipchart and markers in addition to in-house designed templates and gamification cards to stimulate creativity and support groups in the brainwriting process. The brainwriting process involved 4 stages:

- **Stage 1:** Define a “Persona”—either a nurse or a patient with diabetes. The definition must include a short biography, goals, and objectives of the persona, as well as challenges and frustrations.
- **Stage 2:** Describe a typical scenario, either a single encounter for the nurse or a day in the patient’s life, highlighting issues and problems.
- **Stage 3:** Imagine the technologies that can help resolve the problems in the scenario considering the defined characteristics of the persona. Group members then vote for the best resolution.
- **Stage 4:** Rewrite the scenario in stage 2 including the best technology voted for in stage 3.
**RP2: Research Prototype 2**
The design of the second version of the ActiVis RP (RP2) was built on the use cases developed from the first user experiment UX1. Figure 4 shows parts from the different screens of the RP2 separated by a horizontal line while the details of RP2 are presented in the “Results” section.

**UX2.1: Expert Evaluation 1**

**Overview**
An expert review [65] study (UX2.1) of the second prototype (RP2) was conducted at WCQ-M. The expert review study included use cases, surveys, questionnaires, and heuristic walkthroughs. The WCM-Q group were invited for 2 studies.

**UX2.1.1: Study 1**
The participants went through the use cases [66,67] developed by the designers after analysis of the personas and usage scenarios from the first workshop (UX1) conducted with the nursing staff at HMC. Each participant was also asked to follow a think-aloud protocol when performing the task described in the use cases with the RP2 interface. Additional suggestions were provided toward the end of the evaluation in the survey questionnaire. The target users of the use cases were refined based on the suggestions from the participants.
UX2.1.2: Study 2

We used the heuristic walkthrough technique [68] to get participants’ suggestions and improve the prototype further. The participants completed pre- and poststudy questionnaires as well. The identified usability problems were fixed before the updated version of the prototype was further evaluated in the following user experiments (UX2.2).

Participants

The participants were working in the area of diabetes research at WCM-Q. The demographic information of the participants is shown in Table 1.

Table 1. Demographic information of the participants in UX2.1\textsuperscript{a}.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age (years)</th>
<th>Position</th>
<th>Experience (years)</th>
<th>Highest degree or level of school</th>
<th>Competency level in computer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>50-59</td>
<td>Physician</td>
<td>24</td>
<td>Doctorate</td>
<td>Advanced</td>
</tr>
<tr>
<td>Female</td>
<td>30-39</td>
<td>Associate Director, Clinical Research</td>
<td>14</td>
<td>Doctorate</td>
<td>Intermediate</td>
</tr>
<tr>
<td>Male</td>
<td>30-39</td>
<td>Clinical trial Statistician</td>
<td>10</td>
<td>Masters</td>
<td>Advanced</td>
</tr>
</tbody>
</table>

\textsuperscript{a}UX: user study.

Study Protocol

The study protocol used was as follows:

- Participants were invited via email to be a part of the study. In the email, they were informed that the study would be conducted in-person at the campus for their ease.
- On the day of the study, the participants were briefed about the purpose of the study. The participants were informed that notes would be taken during the discussion.
- They were asked to sign a consent form before starting the study. Once signed, they were asked to complete the demographic information as part of the prestudy questionnaire (I2.1).
- The participants were asked to read through the use cases and provide suggestions on how to improve them. For each use case in the questionnaire (I2.2), the participants were asked to choose their most relevant target user, followed by a descriptive comment justifying their choice. The comments would help in making necessary changes to the use cases based on the recommendations when the use case is relevant. Additionally, they were asked 3 closed-ended questions and 1 open-ended question as described in I2.2.
- The participants were asked to evaluate the system using the heuristic walkthrough method [68]. A heuristic walkthrough is an inspection technique that combines the benefits of heuristic evaluations, cognitive walkthroughs, and usability walkthroughs [68]. It is a 2-step process. First, the participants evaluate the system based on a set of tasks and answer questions for each task based on the use cases 1, 2, and 5 from I2.2. Second, the participants identify the usability problems in the prototype and classify them using Nielsen’s heuristics [69] broken down by types of usability issues. The participants were provided a reporting template form (I2.3) to ease the process. Finally, the participants were asked to complete a poststudy questionnaire (I2.4).

Instruments Used

Overview

A total of 4 instruments were used in this study, including a prestudy questionnaire (I2.1), a use case questionnaire (I2.2), a usability problem reporting template (I2.3), and a poststudy questionnaire (I2.4). The details of each instrument and the questions included are provided in the following subsections.

12.1: Prestudy Questionnaire

The prestudy questionnaire gathered basic information on demographic and computer skills from the participants. The questions were about gender, age, job position, university/institution/company (if a student/employed), years of experience, nationality, highest degree, and competency level of the computer.

12.2: Use Case Questionnaire

For each use case, the participants were asked to choose the most relevant target user among 3 possible options, that is, “nurse”, “clinician,” and “not relevant”. The participants were further asked to write a descriptive comment justifying their choice. They were also asked 3 closed-ended questions followed by 1 open-ended question. The participants had to choose the best option based on the 5-point Likert scale (1 for “strongly disagree” to 5 for “strongly agree”) for each close-ended question. The open-ended question was to provide comments for the use case. The closed-ended use case questions (UCQs) are reported in Table 2.
Table 2. Closed-ended questionnaires I2.2 and I2.4.

<table>
<thead>
<tr>
<th>Category and code</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>I2.2 (Use cases)</td>
<td></td>
</tr>
<tr>
<td>UCQ1</td>
<td>It was <em>simple</em> to use this system</td>
</tr>
<tr>
<td>UCQ2</td>
<td>I could effectively <em>complete</em> the tasks using this system</td>
</tr>
<tr>
<td>UCQ3</td>
<td>I was able to complete the tasks <em>quickly</em> using this system</td>
</tr>
<tr>
<td>I2.4 (Overall system)</td>
<td></td>
</tr>
<tr>
<td>OSQ1</td>
<td>Overall, it was <em>easy</em> to use this system</td>
</tr>
<tr>
<td>OSQ2</td>
<td>It was <em>simple</em> to use this system</td>
</tr>
<tr>
<td>I2.4 (Usability)</td>
<td></td>
</tr>
<tr>
<td>USBQ1</td>
<td>It was <em>easy</em> to learn to use this system</td>
</tr>
<tr>
<td>USBQ2</td>
<td>The information provided with this system was <em>clear</em> and easy to understand starting from a search query, navigating by tree keyword levels, up to getting a website description with a link to the targeted website</td>
</tr>
<tr>
<td>USBQ3</td>
<td>It was <em>easy</em> to <em>find</em> the information I needed</td>
</tr>
<tr>
<td>USBQ4</td>
<td>The information was effective in helping me <em>complete</em> the tasks</td>
</tr>
<tr>
<td>USBQ5</td>
<td>The <em>organization</em> of information on the system screens was <em>clear</em></td>
</tr>
<tr>
<td>USBQ6</td>
<td>I <em>liked</em> using the interface of this system</td>
</tr>
<tr>
<td>I2.4 (Usefulness)</td>
<td></td>
</tr>
<tr>
<td>USFQ1</td>
<td>This system has all the functions and capabilities I expect it to have, and</td>
</tr>
<tr>
<td>USFQ2</td>
<td>Overall, I am satisfied with this system performance</td>
</tr>
</tbody>
</table>

*a UCQ: use case question.
*b OSQ: overall system question.
*c USB: usability question.
*d USF: usefulness question.

I2.3: Usability Problems Reporting Template

The template provided the participants with an opportunity to report usability problems that need to be fixed in the prototype. For each usability problem, they were asked to provide a solution/recommendation from their perspective. They were also asked to add a severity rating of the problem as 0 for no problem, 1 for cosmetic, 2 for minor, 3 for major, and lastly 4 for catastrophe.

I2.4: Poststudy Questionnaire

The questionnaire contained 2 closed-ended and 1 open-ended question about the overall system usage, 6 closed-ended questions for usability, and 2 closed-ended questions on the usefulness of ActiVis. For the closed-ended questions, participants had to choose 1 option based on the 5-point Likert scale (1 for “strongly disagree” to 5 for “strongly agree”). The closed-ended questions in the 3 said categories along with the codes assigned to each question are shown in Table 2.

UX2.2: Expert Evaluation 2

Overview

A family doctor was invited to evaluate the second prototype (RP2) to realize the tasks of use cases corresponding to that role from the list refined in UX2.1.1. We followed a subset of the protocol used in UX2.1.2.

Participant

The study involved a Spanish family doctor visiting Qatar Computing Research Institute (QCRI) during October 2018, as part of his collaboration with a former investigator on this project to give feedback on QCRI’s ongoing research projects in the area of medical/health informatics. This physician was from southern Spain where a large proportion of the population are migrants from the Middle East and North Africa (ie, having Arabic origins).

Study Protocol

- The family physician was contacted through email. The participant was invited to take part in the study to share his experience and knowledge, and give feedback on the ActiVis user interface based on 3 use cases refined after UX2.1 that corresponded to the family doctor role (use cases 1, 2, and 5 were selected in Table 3). The participant acknowledged and agreed to be part of the study.
- During the study, the participant was briefed about the purpose of conducting this research and its objectives, and then introduced to the ActiVis user interface. The participant was allowed to have an informal discussion with the researcher to resolve any issues or seek any clarification before they begin the study. Written consent was also taken to be part of the evaluation.
- The participant was informed that notes would be taken throughout the study, the discussion would be
audio-recorded, the interaction during the user interface walkthrough of ActiVis would be recorded through a screen recorder application for the analysis as a backup if any point is missed while taking notes.

- The participant was informed to use a think-aloud protocol while exploring ActiVis based on the use cases. This allowed them to say out loud whatever they were thinking about how to perform a task described in each use case on ActiVis.

Table 3. Use cases (UX2.1).

<table>
<thead>
<tr>
<th>Use case</th>
<th>Initial description resulting from the analysis of UX1* by the designers and evaluated in UX2.1.1</th>
<th>Target user resulting from UX2.1.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use case 1 (Check activity level of a patient): Nurse is at her office; she gets an alert regarding patient sleep quality. Nurse accesses data of the patient; she visualizes the sleep pattern over consecutive days to check how regular it is. She detects irregular sleep time and duration with additional naps on certain days. In particular, she discovers the sleep duration is often short, and the quality of sleep is often poor. She also discovers patient activity is low to moderate.</td>
<td>Nurse/family doctor/clinician researcher</td>
<td></td>
</tr>
<tr>
<td>Use case 2 (Comparing activity between weekdays and weekends): Nurse wants to compare the average activity of the patient across weekdays and on weekends. She wants to identify irregular sleep patterns that could cause more fatigue. She discovers longer sleep duration during weekends. Also, notes that naps mostly occur around 4 PM during weekdays and around 12 PM during weekends.</td>
<td>Family doctor/clinician researcher</td>
<td></td>
</tr>
<tr>
<td>Use case 3 (Comparing 1 individual before and after intervention): Nurse compares the average activity of the patient at different periods, before and after the intervention, to assess the effectiveness of the intervention. She can see the more regular sleep pattern both during weekdays and weekends after the intervention than before it. She can also compare biometrics such as the normalized BMI and weight, between the 2 periods, and she can identify a loss of weight and decrease of BMI.</td>
<td>Family doctor/clinician researcher</td>
<td></td>
</tr>
<tr>
<td>Use case 4 (Comparing 2 individuals [siblings] over a long period): The nurse wants to compare the body metrics and sleep quality of Patient 1 aged 8 years and Patient 2 aged 10 years who are siblings, over a long period to detect a potential family lifestyle issue. The nurse compares the average activity level on weekdays and weekends, and BMI of Patient 1 and Patient 2. She observes that both follow a similar but abnormal pattern of BMI consistent with the average activity level of the corresponding periods, leading to the conclusion that it is a family lifestyle issue.</td>
<td>Clinician researcher</td>
<td></td>
</tr>
<tr>
<td>Use case 5 (Comparing an individual to a group): Nurse compares the average level of activity of the patient with the peer group of the same gender. She can see that the patient is among the overweight subgroup, although her average activity level is similar to one of the normal subgroups, leading her to conclude that the patient may have an unbalanced diet or another health issue affecting her weight.</td>
<td>Family doctor/clinician researcher</td>
<td></td>
</tr>
<tr>
<td>Use case 6 (Comparing males and females of a group before and after intervention): Nurse compares the average level of activity of 2 subgroups of different genders from a group before and after intervention to assess the effectiveness of the intervention. She can see that males increase their activity level after school during weekdays, while females increased their sleep quality, having a more stable bedtime, especially during weekends. She can also compare biometrics such as the normalized BMI and weight, between the 2 periods and she can identify a loss of weight and decrease of BMI more important for the male group.</td>
<td>Clinician researcher</td>
<td></td>
</tr>
</tbody>
</table>

*UX: user study.

Instruments Used

This expert evaluation study used 2 of the instruments (I2.1 and I2.2) described in UX2.1.2.

UX2.3: Users’ Evaluation Workshop 2 With Nurses

Overview

The methods used to conduct this workshop were the same as for UX1. This workshop was conducted with the nursing informatics staff at HMC-Q to evaluate the second prototype (RP2). The workshop was also conducted with the same department and at the same venue as in UX1. It was expected that some of the staff would be the same who attended the first workshop.

The purpose of conducting this workshop was to perform a summative evaluation of the latest version of the prototype and compare it with the Fitbit Dashboard, gather their qualitative feedback, and further improve the user interface.

Participants

The recruitment process of the nursing staff was the same as for UX1. A total of 45 participants, including nurses as well as nursing informatics professionals, attended the workshop.

Study Protocol

The staff of the nursing informatics department was assigned at random to 1 of 4 tables, where each table could accommodate a maximum of 10 participants. Two groups were randomly chosen and assigned to work with the Fitbit Dashboard, while the remaining 2 groups were assigned to work with the ActiVis Dashboard. All the groups were provided a laptop to explore the assigned dashboard in a web browser using temporary credentials to log-on to the dashboard. Each group was instructed to appoint 1 participant as a group representative who would lead the evaluation and inform them about the tasks to be performed. Each group was also instructed to nominate 1 participant as a group secretary who would document the entire discussion and problem found as a part of the evaluation. Each group was also given a task-driven walkthrough template.
Instruments Used

Two instruments were used in this study. These include (1) task-driven walkthrough template, and (2) heuristic evaluation of the dashboard (RP2). The details of each instrument and the questions included are provided in the following subsections. Heuristic evaluation is a usability inspection method that uses evaluators to identify and assess the usability problems in a user interface as a part of the iterative design process. This method relies on the expertise of the domain experts to identify the usability problem in a user interface that needs to be fixed, categorize each identified problem in the heuristics, and rate its severity. The set of 10 heuristics by Nielsen [69] (Figure 5) is the most commonly used in the industry.

Figure 5. UX Check chrome extension [70] showing Nielsen’s 10 heuristics [69]. UX: User study.

![UX Check](image-url)
I4.1: Task-Driven Walkthrough Template

The template contained the following 3 tasks. These 3 tasks were derived from use case 1 (Table 3) proposed after analysis by the designers of the results of the collaborative workshop with nurses (UX1) and validated as a result of UX2.1 with clinicians. Use case 1 is targeted specifically at nurses. These task numbers would be referred to in the results of UX2.3.

- Task 1: Search for the average number of steps for last week.
- Task 2: Search for average active minutes for last month.
- Task 3: Search and describe sleep patterns from May 20 to July 31, 2015.

Each group was asked to brainstorm about the steps needed to complete the task. To guide on how to come up with concrete steps, the following steps were required to complete the first task.

- Enter Patient’s Name/Search in dropdown
- Navigate to Charts
- Observe the particular chart

For each step, the group was asked to answer the following questions:

- Will the user realistically be trying to do this action?
- Is the action visible?
- Will the user recognize the action as being the correct one?
- Will the user understand the feedback/Is the feedback appropriate?

I4.2: Heuristic Evaluation of the Dashboard

For the heuristic evaluation, each group was instructed to download and add the “UX Check” [70] extension in the Google Chrome browser. This extension allows an interactive way to identify and describe the usability problems found on the web page. Opening the extension while staying on any page will show the UX Check panel on the left side of the browser as shown in Figure 5.

The extension will create the necessary regions that can be selected using a single click of the mouse. Users first need to identify any region that contains the usability problem. Clicking on the region will pop-up the dialog as shown in Figure 6. The pop-up allows users to add the heuristic problem, problem description in the form of notes, possible recommendations to fix the problem from their perspective, and lastly the severity rating. The numbers and associated description of the rating are discussed in the “Results” section. Users can save the problem for reporting or cancel their actions. The extension provides a facility for users to view all the identified problems by clicking on the “View progress” link in the pop-up shown on the left side of the web browser. They can export all the problems identified to a Microsoft Word Document by clicking on the “Export” link.

![Figure 6. Problems description and recommendation with UX Check [70].](image)

Ethics Approval

The ethical approval was sought from the Qatar Biomedical Research Institute Institutional Review Board of Hamad Bin Khalifa University, Qatar, before conducting this research (QBRI-IRB 2018-019). The health care professionals as potential users were involved in all the studies as part of this research. Following the cycles of user-centered design, each study on a prototype with health care professionals provided feedback, which was used as a requirement to design an improved version as the next prototype.
Results

**RP1: Visual Analytic Tool for Actigraphy Sensor Data**

We developed 2 versions of the ActiVis interface. The first version (RP1) is shown in Figure 7 and was used in UX1. It was the result of the previous analysis not reported in this study. We proposed a visualization focused on 2 generic tasks: patient overview and comparison, inspired from the discussion with a previous “obesity camp” project participants, and based on the available data [62,63].

Data are body metrics (eg, BMI, weight, height) measured at regular intervals during the obesity camp, together with minute-based activity recordings from wearable accelerometers.

The interface supports an overview and comparison between the data of 2 patients, or 1 patient and a group of patients. The left panel allows selecting the patient and the body metrics features to be displayed. The right panel shows multiple line charts coding for each of the selected features through time coded on the horizontal axis. Color of the line (orange or purple) represents the selected patient or group (Figure 8). The top and bottom rows show bar charts representing the breakdown of activity levels averaged per day for the corresponding patient or group (orange or purple color of the frame; see details in Figure 9). The rightmost views show bar charts averaging the activity level per hour across the selected time window, during weekdays (first and fourth rows) and weekend days (second and third rows). The selection is done by a range selection on the central bar charts and all charts are cross-linked to focus on the same period.

**Figure 7.** First version (RP1) of the ActiVis tool: the left panel is used for patient and group data selection, and filtering on body metrics and activity features; the right panel shows the resulting display for overview and patient/patient and patient/group comparison.

![Figure 7](image_url)

**Figure 8.** Details of the line chart: this chart shows the evolution of the body metric of interest (vertical axis) through time (horizontal axis) for a single patient (blue line), and a group of patients showing its minimum (orange bottom line), maximum (orange top line), and average (red line) values.

![Figure 8](image_url)
UX1: Users’ Evaluation Workshop 1 With Nurses

Nurses have various goals, challenges, and frustrations; however, the results showed that they are mainly concerned about patients’ awareness of their health condition and ways to monitor patients between visits. They particularly need to keep track of patients’ metrics, activity levels, and dietary habits so that they can contact the patients to guide them or remind them about what they have to do as per their activity prescription. Regarding the use of technology, some nurses raised literacy issues and others highlighted accessibility and security concerns.

Nurses highlighted that mobile health (mHealth) apps are an effective means to influence patients’ lifestyles. The most desirable functionalities are activity tracking, dietary advice, and patient education. Including a chat service to facilitate patient-nurse communication is also a viable functionality. Social networking with family and friends is crucial to encourage patients to improve their lifestyles. Interactivity features such as gamification and rewarding achievements were identified as potential ways to motivate patients. Enabling interaction with the app and eliciting patients’ feedback facilitate tailoring contents to suit patient needs.

Outcomes of the workshop showed that recent developments in mHealth apps meet the needs and expectations of their potential users. This is consistent with the latest research findings that confirmed the popularity of mHealth apps (eg, [36]).

The analysis of the workshop usage scenarios led us to design 6 use cases reported in the left-side column of Table 3.

RP2: Research Prototype 2

Figures 10-12 show the resulting interface to support the use cases detailed in Table 3. The interface now has 3 different views to support detailed activity analysis of a patient (use case 1) in Figure 10, qualitative comparison of average activities between patients and groups of patients (use cases 2-6) in Figure 11, and quantitative analysis of the same cases in Figure 12.

Figure 10. This view supports Use Case 1: Check activity level of a patient. It shows the weekly activity of a patient broken down by day. Each row is a day, and the x-axis shows the hours from noon to noon to focus on weekly patterns of sleep (blue). The user can switch the view (top radio button) to span from midnight-to-midnight range and focus on daily activity level (reddish color). This view gives more details of each day and night, allows a side-by-side comparison, and supports the user in detecting activity patterns across several days.
Figure 11. This view supports the qualitative pattern analysis described in Use Cases 2, 3, 4, 5, and 6. It shows a filter (top) to enable the comparison of average weekly activity between a patient or a group (left column) to another patient or a reference group (right).
UX2.1: Expert Evaluation 1

UX2.1.1: Study 1: Use Case Questionnaire

None of the participants selected an option of “Not relevant,” so all the use cases were retained and modified based on the participants’ recommendations.

Table 3 presents 6 use cases resulting from our analysis of UX1, and their reassignment to the correct target user based on the feedback of the participants in UX2.1.1.

The discussion of the results with the participants led us to further distinguish between nurses (use case 1), family doctors (use cases 1, 2, 3, and 5), and clinician researchers (all use cases) types of users. Indeed, the role of a nurse is to observe that a prescribed activity level is correctly followed by the patients to give them reminders if needed, and to notice possible anomalies to report to the doctor, both tasks falling under use case 1. The role of a family doctor is to recommend treatment to the patient. In addition to realizing the tasks assigned to a nurse, the doctor can compare activities or biometrics of a patient between 2 periods (use case 2) to spot differences and recommend a corrective intervention to the patient. The doctors can also control the effect of their prescribed intervention by comparing activity levels and other biometrics before and after it took place (use case 3). At last, the family doctor can compare the patient with statistics derived from groups of patients with similar attributes (age, gender, BMI, or health condition; use case 5). Both nurses and doctors are focused on a single patient at a time. Finally, the clinician researcher focuses on observing trends and patterns within and between cohorts of patients (use case 6), generating knowledge that can guide the family doctors to address the health issue of a specific patient. The clinician researcher can also study more specific cases comparing them over a long period (use case 4) and in general conduct all the other tasks assigned to doctors and nurses for specific patients.

Figure 12. This view supports the quantitative pattern analysis described in Use Cases 2, 3, 4, 5, and 6. It shows the quantitative distribution of two groups of patients along different dimensions as histograms (top two rows) or combined as a color-coded scatterplot (bottom row).
Following this refined assignment, we selected use cases 1, 2, and 5 for further summative evaluation in UX2.1.2 and UX2.2 with the family doctor, because use cases 2 and 3 involved similar tasks. We also focused strictly on use case 1 for the summative evaluation with nurses in UX2.3 as it was the only use case targeted to them.

Table 4 presents the cumulative responses related to the use case questionnaire (I2.2). The first column presents the 3 use cases used in the expert evaluation study (see Table 3 for the full forms of mentioned use cases), the short-form of 3 questions asked for each use case is presented in the subcolumn (see I2.2 and Table 2 for the full form of each question), while the remaining columns contain the cumulative responses in terms of “strongly disagree”, “disagree”, “neutral”, “agree”, and “strongly agree.” Because 3 participants took part in the study, the maximum number of responses is less than or equal to 3. For each use case, the participants found that they were able to effectively, quickly, and efficiently complete the tasks using RP2.

Table 4. Cumulative responses of the participants for use cases 1, 2, and 5.

<table>
<thead>
<tr>
<th>Use case and usability criterion</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>UCQ1a</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>UCQ2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>UCQ3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>UCQ1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>UCQ2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>UCQ3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>UCQ1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>UCQ2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>UCQ3</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

*UCQ: user case question.

**UX2.1.2: Study 2**

**Usability Problems Reporting**

The descriptive comments provided by the participants as part of open-ended questions are presented in Multimedia Appendix 1. It is to be noted that minor changes were incorporated in the user interface of RP2 based on the participants’ comments; therefore, no new RP was produced.

**Poststudy Questionnaire**

Table 5 presents the cumulative responses of the overall system, usability, and usefulness from the participant’s point of view as a part of the poststudy questionnaire (see I2.4 for the questions based on the codes used in the subcolumn) using a “clustered column chart.” The format of Table 5 is similar to that of Table 4.

Table 5 shows the usefulness of the system from the participants’ point of view. The participants found that the system had all the functions and capabilities they expected it to have, and they were satisfied with the performance of this system.

In terms of the overall system, Table 5 shows that the participants found that the system was easy and simple to use.

In terms of the usability of the system, Table 5 shows that the participants found that the system was easy to learn, the information provided was clear and easy to understand, the information needed was easy to find, information was effective to complete the tasks, organization of information across the screens was clear, and lastly, they liked using the interface of this system.
Table 5. Cumulative responses of the participants UX2.1.2.

<table>
<thead>
<tr>
<th>Category and code</th>
<th>Strongly disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overall system</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSQ1(^a)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>OSQ2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td><strong>Usability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USBQ1(^b)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>USBQ2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>USBQ3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>USBQ4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>USBQ5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>USBQ6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Usefulness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USFQ1(^c)</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>USFQ2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

\(^a\)OSQ: overall system question.
\(^b\)USB: usability question.
\(^c\)USF: usefulness question.

UX2.2: Expert Evaluation 2
The audio-taped RP2 interface walkthrough was analyzed. The problems identified and the recommendations provided by the participant evaluation based on the given use cases are presented in Multimedia Appendix 2. Both problems and recommendations were communicated to the engineers to incorporate necessary changes in ActiVis RP2, leading to minor changes in the user interface of ActiVis RP2 used for the UX2.3.

UX2.3: Users’ Evaluation Workshop 2 With Nurses

**Fitbit Dashboard**

**Overview**
Table 6 shows the cumulative number of “yes” and “no” against each question for all the steps required to complete tasks 1, 2, and 3 (see I4.1 for the task details) by all the groups using the Fitbit Dashboard. If the answer to any question is “yes,” then it means the group mutually agreed to the statement; however, if an answer to any question is “no,” then it shows the disagreement. In the latter case, they were instructed to add more description so that the problem can be rectified in the user interface. However, during the analysis of the filled templates returned by the groups, it was found that some of the groups also commented when their answer was “yes.” Such comments mainly reflected the minor changes recommended by the group despite an agreement to the question.

Table 6. Cumulative number of responses against each question for all the steps required to complete tasks using the Fitbit Dashboard.

<table>
<thead>
<tr>
<th>Fitbit</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q1: Will the user realistically be trying to do this action?</td>
</tr>
<tr>
<td></td>
<td>Q2: Is the action visible?</td>
</tr>
<tr>
<td></td>
<td>Q3: Will user recognize the action as being the correct one?</td>
</tr>
<tr>
<td></td>
<td>Q4: Will the user understand the feedback/is the feedback appropriate?</td>
</tr>
<tr>
<td>Task 1:</td>
<td>Yes</td>
</tr>
<tr>
<td>6 steps</td>
<td>6</td>
</tr>
<tr>
<td>Task 2:</td>
<td>4</td>
</tr>
<tr>
<td>4 steps</td>
<td>4</td>
</tr>
</tbody>
</table>

The results for each task are as follows:

**Task 1**
For all the steps in Q1, the participants were willing to perform an action. For most of the steps (5/6) in Q2 and Q3, the participants found that the action was visible, and they could recognize that the action performed was the correct one. For 4/6 steps in Q4, the participants found that they were able to understand the feedback, or that the feedback was appropriate.
Tasks 2 and 3
For all the steps (4/4), the participants were willing to perform an action, found that the action was visible, that they recognized that the action performed was the correct one, and that the feedback given toward the end of the task was understandable or appropriate.

ActiVis Dashboard
Overview
Table 7 shows the cumulative number of “yes” and “no” against each question for all the steps required to complete tasks 1, 2, and 3 by all the groups using the ActiVis Dashboard.

The format of Table 7 is similar to that of Table 6. The results for each task are as described in the following sections.

| Table 7. Cumulative number of responses against each question for all the steps required to complete tasks using the ActiVis Dashboard. |
|-------|----------------|----------------|----------------|----------------|
|       | Questions                  | ActiVis | Q1: Will the user realistically be trying to do this action? | Q2: Is the action visible? | Q3: Will user recognize the action as being the correct one? | Q4: Will the user understand the feedback/Is the feedback appropriate? |
|       |                             | Yes   | No | Yes | No | Yes | No | Yes | No |
| Task 1: 6 steps | Q1: Yes | 5 | 1 | 6 | 0 | 4 | 2 | 3 | 3 |
| Task 2: 6 steps | Q1: Yes | 5 | 1 | 3 | 3 | 4 | 2 | 4 | 2 |
| Task 3: 6 steps | Q1: Yes | 4 | 2 | 5 | 1 | 5 | 1 | 6 | 0 |

Task 1
For most of the steps (5/6) in Q1, the participants were willing to perform an action, for all the steps (6/6) in Q2, the participants found that the action was visible. For 4/6 steps in Q3, the participants were able to recognize that the action performed was the correct one. However, for 3/6 steps in Q4, the participants had mixed opinions; for half of the steps, they found that they were either unable to understand the feedback, or that the feedback was inappropriate, while for the remaining steps, they found that they were able to understand the feedback, or that the feedback was appropriate.

Task 2
For most of the steps (3/4) in Q1, the participants were willing to perform an action; however, for 3/5 steps in Q2, the participants found that the action was not visible. For 2/4 steps in Q3, the participants had mixed opinions. For half of the steps, some participants found that they were able to recognize the action performed, while the other participants found that they were unable to recognize the action performed. Similarly, a mixed opinion was also found for Q4 (2/4 steps). For half of the steps, some participants found that they were able to understand the feedback given toward the end of the task, while the other participants found that they were unable to understand the feedback given at the end of the task.

Task 3
For 3/5 steps in Q1, the participants were willing to perform an action, for 4/5 steps in Q2 and Q3 each, the participants found that the action was visible and that they recognized that action performed was the correct one. For all the steps, the participants found that the feedback given after the task was performed was understandable or appropriate.

Heuristic Evaluation of the Interfaces
Figure 13 shows the number of usability problems found and the average severity ratings of the identified problems in the Fitbit Dashboard and the ActiVis Dashboard, respectively, using Nielsen’s 10 heuristics. The “stacked columns” represent the “number of usability problems” (left vertical scale), whereas the “line with markers” represents the “average severity rating of the identified problems” (right vertical scale). Each stack column shows the number of usability problems found based on the 4 severity ratings, that is, cosmetic, minor, major, and catastrophic. The axis on the left-hand side is known as the primary axis and it is related to the “stacked columns,” whereas the axis on the right-hand side is known as the secondary axis and is related to the “line with markers.”

A total of 11 usability problems were identified in each of the 2 dashboards (ie, Fitbit and ActiVis). The analysis of the results in terms of the number of usability problems found in Fitbit shows that the recognition heuristic (n=4) was the more commonly broken heuristic, followed by the visibility and control heuristics (n=2 each). Similarly, the analysis of the results in terms of the number of usability problems found in ActiVis shows that the control heuristic (n=4) was the more commonly broken heuristic, followed by the visibility, match, and recognition heuristics (n=2 each).

The analysis of the results in terms of the average severity rating shows that the majority of problems identified are minor. The number of usability problems identified and their severity rating provided by the participants for the Fitbit Dashboard and the ActiVis Dashboard were the same. However, the Fitbit Dashboard has more severe issues than the ActiVis Dashboard in terms of visibility, recognition, error, and documentation. Still, ActiVis needs improvement compared with Fitbit in terms of control and match, and to solve the catastrophic visibility issue identified.
RP3: Research Prototype 3

The work on this project is still ongoing. The 3 UX (UX2.1, UX2.2, and UX2.3) of the second ActiVis prototype (RP2) led to new and updated requirements for RP3. Since the last study, the work on this interface has been organized in 2 different branches. The research effort specific to the visualization interface has been split between the different types of users (nurses described in use case 1; family doctors in use cases 2, 3, and 5; and clinician researchers in use cases 4 and 6) with specific charts and interactions but with a common core of data processing functions. The developed visualization prototypes are planned to be integrated into a platform able to read data from different wearable devices available on the market, and integrated into a clinic environment. User evaluations will continue as part of the user-centered design and PD cycles.

Discussion

Principal Findings

The key finding from these PD studies is the derivation from post hoc analysis of nurses’ workshop, and the validation by 2 physicians, 1 clinician researcher, and 1 clinician statisticians of the 6 use cases to analyze wearable data for health care professionals. These use cases are assigned to specific user roles: nurses, family doctors, and clinician researchers. They will facilitate the design and development of new data analytics and visualization interfaces to support the particular needs of these users.

UX1

During the PD workshop with nurses evoking their work and relations with patients and other health care professionals, we could not identify specific cultural needs in terms of the visualization of wearable data for health care professionals. Some of the persona and usage scenarios were obviously representative of the local Arabic culture by design, and it is also well-known that particular customs such as prayer times and fasting during the Ramadan Holy month can impact people’s patterns of physical activities, sleep, and diet, but none of these aspects finally influenced the more technical use cases we derived from these discussions. The use cases we propose ended up being culturally agnostic (Table 3).

UX2.3

The final evaluation comparing Fitbit and ActiVis dashboards showed there is ample room for improvement even in existing interfaces such as Fitbit, widely available for the general public. We only evaluated use case 1 specific to nurses and already identified some major and catastrophic problems, with severe ratings being more frequent with Fitbit than with ActiVis. Although Fitbit was not necessarily designed to support this use case, it shows that we cannot simply reuse available interfaces to support end users in the best way. Supporting statistical and visual analyses of wearable data from cohorts of patients as stated in use cases targeted at clinician researchers are not optimal or even possible with existing visualization tools and will deserve further investigations.

In general, this project also showed how conducting PD is necessary but still challenging. It has been difficult to plan several of the studies in advance. The use of the opportunistic approach allowed us to use the available local health care professionals throughout the design, development, and validation of RPs presented in this paper. Qatar is a country where 90% of the population are expatriates mixing Western, Asian, and Muslim cultures. Because of the heterogeneous culture and origin of the population, it is challenging to study the levels of health awareness in Qatar [71]. Nevertheless, this is crucial to understand to develop efficient health-targeted visualizations. The population diversity also allowed us to get feedback from non-Qatari, non-Muslim users too. Opening to a wide range of cultures in the same place is of interest to understand what is common or specific to these end users. Although the interface for health care professionals is not impacted by local culture, we know from a previous study [72] that the interfaces involving the patients themselves will need specific care of their local particular health conditions (eg, diabetes or obesity) and Muslim culture (Ramadan Holy month effect on diet, sleep, and physical activity).

Limitations

The study has several limitations. First, a specific set of methods from the user-centered design and PD methodologies was used. Second, the studies were conducted with a selected list of institutions and their experts as participants. Third, several participants were used in each study that was mainly dependent on multiple factors, including availability based on their routine clinical appointments, meetings, and teaching. Fourth, Nielsen’s
heuristics were used to diagnose user problems in the prototype that need to be fixed. All these constraints could affect the generalizability of the results. For future studies, we seek higher diversity and a higher number of participants, and domain-specific heuristics to get more generalizable findings.

Conclusion
This paper shows how the use of PD and user-centered design allowed the development of a visualization interface supporting the real needs of health care professionals in Qatar. Although Qatar is an oil-based economy that nurtures a rich multicultural environment, the use cases we derived from the PD studies happen to be culturally agnostic. We hope these use cases will serve to design future visualization and analytic systems optimized to support the needs specific to nurses, family doctors, and clinician researchers, beyond existing dashboards designed primarily for the general public. This work is still ongoing. A cluster project has now started that is funded under the Qatar National Research Fund [73] and will support further development and integration of these visualizations in a clinical setting to help clinician researchers, doctors, and nurses improve the health of Qatari citizens and residents.

Acknowledgments
We thank Dr Luis Fernandez-Luque for his strong support and thoughtful guidance throughout this research.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Expert comments.
[DOCX File, 23 KB - humanfactors_v9i2e25880_app1.docx ]

Multimedia Appendix 2
Expert recommendations.
[DOCX File, 24 KB - humanfactors_v9i2e25880_app2.docx ]

References


70. Gallello C. UX Check. URL: https://www.uxcheck.co/ [accessed 2022-11-20]


**Abbreviations**

- **HMC-Q**: Hamad Medical Corporation-Qatar
- **PD**: participatory design
- **QCRI**: Qatar Computing Research Institute
- **RP**: research prototype
- **UX**: user study
- **WCM-Q**: Weill Cornell Medicine - Qatar
- **WHO**: World Health Organization
Corrigenda and Addenda

Correction: Improving Pelvic Floor Muscle Training Adherence Among Pregnant Women: Validation Study

Aida Jaffar1,2, MMed; Sherina Mohd-Sidik1, PhD; Chai Nien Foo3, PhD; Novia Admodisastro4, PhD; Sobihatun Nur Abdul Salam5, PhD; Noor Diana Ismail6, MMed

1Department of Psychiatry, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Selangor, Malaysia
2Primary Care Unit, Faculty of Medicine and Defence Health, Universiti Pertahanan Nasional Malaysia, Wilayah Persekutuan, Malaysia
3Department of Population Medicine, Universiti Tunku Abdul Rahman, Selangor, Malaysia
4Software Engineering & Information System Department, Faculty of Computer Science & Information Technology, Universiti Putra Malaysia, Selangor, Malaysia
5School of Multimedia Technology and Communication, College of Arts and Sciences, Universiti Utara Malaysia, Kedah, Malaysia
6Klinik Kesihatan Bt 9 Cheras, Ministry of Health, Selangor, Malaysia

Corresponding Author:
Sherina Mohd-Sidik, PhD
Department of Psychiatry
Faculty of Medicine and Health Sciences
Universiti Putra Malaysia
Serdang
Selangor, 43400
Malaysia
Phone: 60 3 9769 2541
Fax: 60 3 9769 2706
Email: sherina@upm.edu.my

Related Article:
Correction of: https://humanfactors.jmir.org/2022/1/e30989
doi:10.2196/38175

In “Improving Pelvic Floor Muscle Training Adherence Among Pregnant Women: Validation Study” (JMIR Hum Factors 2022;9(1):e30989), the following errors were noted.

1. Abstract:
In the originally published paper, the first sentence in the abstract was stated as

Mobile health apps, for example, the Tât, have been shown to be potentially effective in improving pelvic floor muscle training (PFMT) among women, but their effectiveness in pregnant women was limited.

This has been corrected to:

Mobile health apps, for example, the Tât, have been shown to be potentially effective in improving pelvic floor muscle training (PFMT) among women, but they have not yet been studied among pregnant women.

2. Methods, Intervention Mapping:
The originally published paper was missing two references for this statement:

The outcomes of the intention are self-efficacy (17 questions) and adherence (6 questions).

This has been corrected to:

The outcomes of the intention are self-efficacy (17 questions) and adherence (6 questions).

3. Methods, Cross-Sectional Study:
The originally published paper was missing two references for this statement:

The findings from this study provided input for the content of their educational videos and short notes on PFMT, which were captured as frequently asked questions (FAQ).

This has been corrected to:

The findings from this study provided input for the content of their educational videos and short notes on PFMT (45, 46) which were captured as frequently asked questions (FAQ).

4. Results:
The originally published paper stated the following in row 1, column 2 of Table 5:

System credibility-expertise and authority.

This has been corrected to:

System credibility-expertise and authority.
2. Primary support—Virtual rehearsal principle
This has been corrected to:
System credibility-expertise and authority

5. Results:
The originally published paper stated the following as the title for the first column of Table 5:

COM-B model and behavioral change techniques incorporated in the mHealth app.

This has been corrected to:

COM-B model and features of the mHealth app.

6. Discussion:
The originally published paper was missing one reference for this statement:

The PSD component of the system’s credibility and trustworthiness, with the expertise involved in the development, may add to the user’s sense of safety and reliability regarding the KEPT app.

This has been corrected to:

The PSD component of the system’s credibility and trustworthiness (55), with the expertise involved in the development, may add to the user’s sense of safety and reliability regarding the KEPT app.

6. References:
In the corrected paper, the following citations have been newly added to the Reference List. As these new references have been numbered per the order of their in-text citations, the remaining citations in the reference list have been renumbered accordingly.


The correction will appear in the online version of the paper on the JMIR Publications website on April 11, 2022, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 21.03.22; this is a non–peer-reviewed article; accepted 23.03.22; published 11.04.22.

Please cite as:
Jaffar A, Mohd-Sidik S, Foo CN, Admodisastro N, Abdul Salam SN, Ismail ND
Correction: Improving Pelvic Floor Muscle Training Adherence Among Pregnant Women: Validation Study
JMIR Hum Factors 2022;9(2):e38175
URL: https://humanfactors.jmir.org/2022/2/e38175
doi:10.2196/38175
PMID:35404829

©Aida Jaffar, Sherina Mohd-Sidik, Chai Nien Foo, Novia Admodisastro, Sobihatun Nur Abdul Salam, Noor Diana Ismail. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 11.04.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Original Paper

Smartphone Alcohol Use Disorder Recovery Apps: Cross-sectional Survey of Behavioral Intention to Use

Rijuta Menon¹, BSc, MSc; Julien Meyer¹, BA, MSc, PhD; Pria Nippak¹, BA, BSc, MSc, PhD; Housne Begum¹, BSc, MSc, PhD
School of Health Services Management, Ted Rogers School of Management, Ryerson University, Toronto, ON, Canada

Corresponding Author:
Julien Meyer, BA, MSc, PhD
School of Health Services Management
Ted Rogers School of Management
Ryerson University
350 Victoria Street
Office: TRS 3-081
Toronto, ON, M5B 2K3
Canada
Phone: 1 4169951338
Email: julien.meyer@ryerson.ca

Abstract

Background: Alcohol use disorder (AUD) carries a huge health and economic cost to society. Effective interventions exist but numerous challenges limit their adoption, especially in a pandemic context. AUD recovery apps (AUDRA) have emerged as a potential complement to in-person interventions. They are easy to access and show promising results in terms of efficacy. However, they rely on individual adoption decisions and remain underused.

Objective: The aim of this survey study is to explore the beliefs that determine the intention to use AUDRA.

Methods: We conducted a cross-sectional survey study of people with AUD. We used the Unified Theory of Acceptance and Use of Technology, which predicts use and behavioral intention to use based on performance expectancy, effort expectancy, social influence, and facilitating conditions. Participants were recruited directly from 2 sources; first, respondents at addiction treatment facilities in Ontario, Canada, were contacted in person, and they filled a paper form; second, members from AUD recovery support groups on social media were contacted and invited to fill an internet-based survey. The survey was conducted between October 2019 and June 2020.

Results: The final sample comprised 159 participants (124 involved in the web-based survey and 35 in the paper-based survey) self-identifying somewhat or very much with AUD. Most participants (n=136, 85.5%) were aware of AUDRA and those participants scored higher on performance expectancy, effort expectancy, and social influence. Overall, the model explains 35.4% of the variance in the behavioral intention to use AUDRA and 11.1% of the variance in use. Social influence (P=.31), especially for women (P=.23), and effort expectancy (P=.25) were key antecedents of behavioral intention. Facilitating conditions were not significant overall but were moderated by age (P=.23), suggesting that it matters for older participants. Performance expectancy did not predict behavioral intention, which is unlike many other technologies but confirms other findings associated with mobile health (mHealth). Open-ended questions suggest that privacy concerns may significantly influence the use of AUDRA.

Conclusions: This study suggests that unlike many other technologies, the adoption of AUDRA is not mainly determined by utilitarian factors such as performance expectancy. Rather, effort expectancy and social influence play a key role in determining the intention to use AUDRA.

(JMIR Hum Factors 2022;9(2):e33493) doi:10.2196/33493

KEYWORDS
mobile health; alcohol use disorder; disease management; mobile apps; Unified Theory of Acceptance and Use of Technology
Introduction

Alcohol causes 3.3 million deaths a year worldwide, close to 6% of all deaths [1]. Many of these deaths are associated with alcohol use disorder (AUD), defined as “a problematic pattern of alcohol use accompanied by clinically significant impairment or distress” [2]. Treatment and engagement with recovery activities, such as brief interventions, motivational enhancements, and cognitive behavior therapies, are integral to avoiding disease progression [3]. They are well accepted and effective. However, they usually require substantial time, money, and resources; moreover, they depend predominantly on the skill of the clinician and can be stigmatizing [3].

With the advent of smartphones, mobile health (mHealth) apps have been developed to address AUD recovery. These apps can provide information and advice on how to address the condition and help users track their behavior. They serve as accessible, widespread, cost-effective, dependable, individualized, and anonymous alternatives or complements to traditional interventions [3]. These apps have also proved invaluable in the context of the COVID-19 pandemic, which has aggravated addiction issues while severely restricting access to in-person support services. In a 2019 literature review on the efficiency of AUD recovery apps (AUDRA), 63% (n=12) of the 19 studies considered found significant evidence of positive outcomes, 32% (n=6) found none, and 5% (n=1) found negative outcomes for some users [1]. Positive outcomes included decreased alcohol consumption, decreased episodes of binge drinking and alcohol-related injuries, and decreased addiction levels. Despite these benefits, evidence from mHealth app studies indeed suggest low adoption rates [4,5], and studies about the acceptance of mental health apps particularly suggest that potential users remain unconvinced of their usefulness.

Methods

Study Design and Survey Instrument

This study is a cross-sectional survey of nonusers or existing users of AUDRA. The survey covered the factors contributing to the behavioral intention to use smartphone AUD recovery apps among participants (it targeted use of AUDRA in general and not of any specific app). The UTAUT framework and model questionnaire items (Figure 1 and Textbox 1) were adapted to measure the constructs, particularly its operationalizations from UTAUT2. UTAUT predicts that the behavioral intention to use a technology depends on four factors: (1) performance expectancy, defined as the degree to which using a technology will provide benefits to consumers in performing certain activities; (2) effort expectancy, defined as the degree of ease associated with consumers’ use of technology; (3) social influence, defined as the extent to which consumers perceive that important others (e.g., family and friends) believe they should use a particular technology; and (4) facilitating conditions, defined as consumers’ perceptions of the resources and support available to perform a behavior [6,7].

Figure 1. Unified Theory of Acceptance and Use of Technology research model showing the complete theoretical model with the moderating relationships [6].

The constructs of hedonic motivation, price value, and habit from UTAUT2 were removed. They are not applicable to this study as AUDRA are not primarily designed for enjoyment; almost all AUDRA are free on app stores, and AUDRA are still new and rare, which diminish the importance of habit and experience. Age and gender also moderate these relations. Figure 1 shows the theoretical model with the moderating relationships. The constructs were measured by adapting the 16 corresponding items from UTAUT2 [7] using a 5-point Likert scale ranging...
from “strongly disagree” to “strongly agree,” except for behavioral intention that had choices “yes,” “no,” or “maybe” and use, which used a 6-point Likert scale ranging from “everyday” to “at least once a year” (Textbox 1). A follow-up survey was conducted 6 months later to investigate the subsequent usage behavior. The study was approved by Ryerson’s Research Ethics Board (approval reference number: 2019-277).

Textbox 1. Survey items used for each construct.

**Performance expectancy**

1. I find/would find Smartphone Alcohol Use Disorder (AUD) recovery apps useful in complementing the daily activities I do to help me recover.  
2. Using Smartphone AUD recovery apps helps/would help me learn recovery skills more quickly.  
3. Using Smartphone AUD recovery apps helps/would help me increase the effectiveness of activities I do to help me recover.

**Effort expectancy**

4. Learning how to use Smartphone AUD recovery apps is/would be easy for me.  
5. My interaction with Smartphone AUD recovery apps is/would be clear and understandable.  
6. I find/would find Smartphone AUD recovery apps addiction recovery apps easy to use.  
7. It is/would be easy for me to become skillful at using Smartphone AUD recovery apps.

**Social influence**

8. People who are important to me think that I should use Smartphone AUD recovery apps.  
9. Caregivers think that I should use Smartphone AUD recovery apps.  
10. People who influence my behavior think that I should use Smartphone AUD recovery apps.  
11. People whose opinions that I value prefer that I use Smartphone AUD recovery apps.

**Facilitating conditions**

12. I have the resources necessary to use Smartphone AUD recovery apps.  
13. I have the knowledge necessary to use Smartphone AUD recovery apps.  
14. Smartphone AUD recovery apps are compatible with other technologies I use.  
15. I can get help from others to use Smartphone AUD recovery apps.

**Behavioral intention**

16. Do you intend to use or keep using a Smartphone AUD recovery app(s)?

**Use**

17. If you are using a Smartphone app that assists with recovery AUD, how often do you use it?

**Recruitment**

Participants were aged 18 years and older, self-identified as having an AUD, and owned a smartphone. Data were collected between October 2019 and June 2020. The survey was offered to participants in 2 modalities. The first was in a pen and paper format, with participants recruited from 6 AUD treatment facilities in Ontario, Canada. Second, an internet-based version of the survey was shared on various English-speaking AUD recovery groups. Participants were offered a gift certificate for their participation. A second follow-up survey was conducted to track usage longitudinally, but it was discarded due to an insufficient response rate. In the partial least squares-structural equation model (PLS-SEM), the minimum sample size is 10 times the number of paths targeting a particular construct. In our study, this means a minimum of 40 respondents [10].

**Statistical Analysis**

Internal validity was evaluated using the Cronbach alpha and composite reliability (CR) [11]. Values for the Cronbach alpha and CR are considered satisfactory if they are between 0.7 and 0.9 [12]. Convergent validity was assessed using the outer loadings of the indicators and the values of the average variance extracted (AVE) [11]. To help establish convergent validity on a construct, the outer loadings should be 0.708 or higher and the AVE value must be 0.5 or higher to indicate that the construct explains more than 50% of the variance of its indicators [11].

The heterotrait-monotrait ratio was used to assess the discriminant validity between constructs. When constructs are conceptually more distinct, as is the case with the constructs of UTAUT, a lower conservative threshold of 0.85 is suggested such that values above this threshold indicate a lack of discriminant validity [11].

The results of the survey were analyzed using SPSS Statistics (version 26; IBM Corporation) and SmartPLS 3 (version 3.2.9; SmartPLS GmBH). SPSS Statistics was used for descriptive statistics and chi-square tests were performed to test the associations between variables and differences in the mean scores for variables; their determinants between the 2 groups were assessed using t tests at a 95% CI. As the focus of this
study was on identifying the antecedents of smartphone AUD recovery app adoption, there were no exclusion criteria in place to exempt the responses of those who did not possess prior knowledge about the existence of these apps. PLS-SEM was used to test the research model (Figure 1) for its reliability, convergent validity, and the discriminant validity of the constructs. The structural model was assessed using $R^2$ and bootstrapping tests were conducted to examine the statistical significance (taken at 95% CI) of the path coefficients [11]. For the PLS algorithms and bootstrapping calculations, missing data were treated with mean value replacement. SmartPLS 3 was used to test the theoretical model.

The open-ended questions aimed to determine why the participants used or did not use AUDRA. The comments were analyzed quantitatively by themes [13]. Although the low rate of response for these questions did not allow for deriving meaningful statistics, it was sufficient to identify some recurring themes.

### Results

#### User Statistics

A total of 1792 surveys were completed. However, most web-based surveys had to be excluded, with 900 excluded for multiple participations, 416 for answering randomly or incompletely, and 317 for not meeting the inclusion criteria (not identifying with AUD or not owning a smartphone). Finally, 159 surveys (124 web-based and 35 paper surveys) could be used.

Table 1 provides the background characteristics of the respondents. The 159 respondents comprised 111 (69.8%) males, 45 (28.3%) females, and 3 (1.9%) individuals who identified themselves as “other” gender. The average age of the respondents was 36 (SD 10.3) years, with a range of 19 to 65 years and mostly between 19 and 39 years (n=117, 73.6%). More than half (n=94, 59.1%) of the participants disclosed their self-identification with AUD as “Very much like me” and the rest (n=65, 40.9%) disclosed it as “Somewhat like me.” In terms of prior awareness of AUDRA, 94 participants answered “Very much like me” and 65 participants mentioned “Somewhat like me;” prior awareness of AUDRA was exhibited by 136 (85.5%) participants.

#### Table 1. Sociodemographic characteristics of the respondents (N=159).

<table>
<thead>
<tr>
<th>Variable</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>111 (69.8)</td>
</tr>
<tr>
<td>Female</td>
<td>45 (28.3)</td>
</tr>
<tr>
<td>Other/undisclosed</td>
<td>3 (1.9)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>19-39</td>
<td>117 (73.6)</td>
</tr>
<tr>
<td>40-65</td>
<td>39 (24.5)</td>
</tr>
<tr>
<td>Undisclosed</td>
<td>3 (1.9)</td>
</tr>
<tr>
<td><strong>Self-identification with AUD</strong></td>
<td></td>
</tr>
<tr>
<td>Very much like me</td>
<td>94 (59.1)</td>
</tr>
<tr>
<td>Somewhat like me</td>
<td>65 (40.9)</td>
</tr>
<tr>
<td><strong>Prior awareness of AUDRA</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>136 (85.5)</td>
</tr>
<tr>
<td>No</td>
<td>23 (14.5)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>159 (100)</td>
</tr>
</tbody>
</table>

**Table 2 describes the reliability and validity of the constructs.**

Internal validity was evaluated using the Cronbach alpha and CR, with the acceptable range falling between 0.6 and 0.7 [12]. The AVE values for all the constructs, except for facilitating conditions, were above 0.5, thereby indicating convergent validity. Note that the first item, FC1, pertaining to facilitating conditions had to be removed because when FC1 was included along with the other items (FC2, FC3, and FC4), the CR value was very low (0.037). After removing FC1 from facilitating conditions, the CR value improved to 0.621. Therefore, 3 items related to facilitating conditions and all items pertaining to the other constructs were retained.

**Reliability and Validity of the Constructs**
For the heterotrait-monotrait ratio, all comparisons were well under the recommended threshold of 0.85 and indicated satisfactory discriminant validity between the constructs (Table 2).

Then we compared the constructs to investigate differences between respondents. We compared respondents who identified “somewhat like me” and “very much like me” with AUD, as shown in Table 3. The only significant difference was that the “very much like me” group found it slightly easier to use AUDRA.

Third, we compared respondents based on their prior awareness of AUDRA (Table 4). Respondents aware of AUDRA scored significantly higher on performance expectancy, effort expectancy, and social influence than respondents who had not.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Cronbach alpha</th>
<th>Average variance extracted</th>
<th>Composite reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance expectancy (PE); loading</td>
<td>.678</td>
<td>0.593</td>
<td>0.812</td>
</tr>
<tr>
<td>PE1; 0.901</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE2; 0.714</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE3; 0.676</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effort expectancy (EE); loading</td>
<td>.685</td>
<td>0.512</td>
<td>0.806</td>
</tr>
<tr>
<td>EE1; 0.794</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE2; 0.650</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE3; 0.759</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EE4; 0.648</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social influence (SI); loading</td>
<td>.766</td>
<td>0.585</td>
<td>0.849</td>
</tr>
<tr>
<td>SI1; 0.720</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI2; 0.749</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI3; 0.764</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SI4; 0.824</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facilitating conditions (FCs); loading</td>
<td>.412</td>
<td>0.407</td>
<td>0.621</td>
</tr>
<tr>
<td>FC2; 0.395</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC3; 0.976</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC4; 0.335</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Level of identification with alcohol use disorder and participants’ mean scores on Unified Theory of Acceptance and Use of Technology constructs (N=159).

<table>
<thead>
<tr>
<th>UTAUT(^a) constructs</th>
<th>Self-identification with AUD(^b)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average out of 5</td>
<td>Very much like me (n=94) Somewhat like me (n=65)</td>
<td></td>
</tr>
<tr>
<td>Performance expectancy</td>
<td>3.9 4.0</td>
<td>.56</td>
</tr>
<tr>
<td>(3 items)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effort expectancy</td>
<td>4.1 3.9</td>
<td>.05(^c)</td>
</tr>
<tr>
<td>(4 items)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social influence</td>
<td>3.7 3.8</td>
<td>.5</td>
</tr>
<tr>
<td>(4 items)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facilitating conditions</td>
<td>4.1 4.1</td>
<td>.8</td>
</tr>
<tr>
<td>(4 items)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral intention</td>
<td>3.5 3.2</td>
<td>.57</td>
</tr>
<tr>
<td>(1 item)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use behavior</td>
<td>2.6 2.6</td>
<td>.93</td>
</tr>
<tr>
<td>(1 item)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)UTAUT: Unified Theory of Acceptance and Use of Technology.

\(^b\)AUD: alcohol use disorder.

\(^c\)The italicized P value is statistically significant.
Table 4. Prior awareness of the existence of smartphone alcohol use disorder recovery apps and participants’ mean scores on Unified Theory of Acceptance and Use of Technology constructs (N=159).

<table>
<thead>
<tr>
<th>UTAUTa constructs</th>
<th>Prior awareness of smartphone AUDRAb</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average out of 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance expectancy (3 items)</td>
<td>Yes (n=136)</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>No (n=23)</td>
<td>3.6</td>
</tr>
<tr>
<td>Effort expectancy (4 items)</td>
<td></td>
<td>.02c</td>
</tr>
<tr>
<td>Social influence (4 items)</td>
<td></td>
<td>.04</td>
</tr>
<tr>
<td>Facilitating conditions (4 items)</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Behavioral intention (1 item)</td>
<td></td>
<td>.45</td>
</tr>
<tr>
<td>Use behavior (1 item)</td>
<td></td>
<td>.74</td>
</tr>
</tbody>
</table>

aUTAUT: Unified Theory of Acceptance and Use of Technology.
bAUDRA: alcohol use disorder recovery apps.
cThe italicized P value is statistically significant.

Structural Model to Identify the Behavioral Factors

To analyze the model fit, PLS-SEM was used. Figure 2 shows the path coefficients and the statistical significance of the relationships along with the coefficient of determination or the $R^2$ value.

Effort expectancy and social influence were significant predictors of behavioral intention to use smartphone AUDRA, which itself predicted use. However, performance expectancy had no effect on behavioral intention. Gender moderated the effect of social influence, meaning that the effect of social influence on behavioral intention was more significant in women than in men. Facilitating conditions had no significant effect on use except for older users who were more likely to be influenced by facilitating conditions. Overall, the model explains 35.4% of the variance in behavioral intention and 11.1% of the variance in use behavior.

![Figure 2. Complete model showing path coefficients and $R^2$. Statistical significance of the relationships (path coefficients): *P<.05; **P<.01; ***P<.001. AUD: alcohol use disorder; PE: performance expectancy; EE: effort expectancy; SI: social influence; FC: facilitating conditions.](image)

Open-Ended Questions Regarding AUDRA

Open-ended responses provided further insight into participants’ attitudes to AUDRA. Response rates on the 3 questions were between 35% and 67%. Privacy and security concerns were the most frequently given reasons by participants for not wanting to use AUDRA. One respondent stated that “Privacy would be the only issue regarding using an app to help in recovery,” whereas another pointed out “the potential of data tracking and possibility of using my information for profit.” Other frequently given responses pointed to how “confusing” or “complicated” apps could be. Respondents also expressed their skepticism...
over the efficacy of such apps in helping them with AUD recovery and noted specific user-unfriendly features, such as too many reminders, notifications, or advertisements: “Pop-ups asking me to rate and/or buy a pro version. Unsolicited communications.” The participants were also dissuaded from potential AUDRA use if there were technical glitches, or “bugginess,” with the apps.

In terms of what would make them want to use AUDRA, respondents asked if these apps would help them with abstinence and prevent relapse. Users often mentioned how a tracking feature (“track my days [without alcohol] and money savings”) helped them. On the contrary, many other users complained about the lack of a tracking feature in the apps they were using. Respondents also frequently cited the ability of apps to connect them with others through social networking features and with local resources, such as if they could “find a meeting close by” and “…Access to events happening through local AA chapter.” as major reasons why they would be encouraged to use the app.

Discussion

Principal Results

This study investigated the key antecedents of behavioral intention to use AUDRA among people with AUD. Generally, most of the 159 participants (n=136, 85.5%) were aware of AUDRA. This study confirms the role of effort expectancy and social influence as significant predictors of the intention to use AUDRA, similar to the findings of previous UTAUT studies on mHealth [6,14]. This was confirmed by open-ended answers suggesting that some of the main hurdles to use are technical glitches. However, performance expectancy was not found to significantly predict the intention to use from the final model. This is intriguing because this factor is considered the key determinant of technology usage in general [15-18]. However, it does not appear to apply to mHealth apps [14,19-21]. Other studies have highlighted that despite playing a major role, performance expectancy may not prove salient for mHealth [6,14]. This was confirmed by open-ended answers suggesting that some of the main hurdles to use are technical glitches. However, performance expectancy was not found to significantly predict the intention to use from the final model.

Facilitating conditions had no direct effect on use, but they were moderated by age. This suggests that facilitating conditions play a more important role as participants age. Other studies conducted with people aged over 60 [20] and 65 years [20,22] have also found a significant influence of facilitating conditions on the use of mHealth apps. Considering that our sample only had 1 participant aged over 60 years, this suggests that the importance of facilitating conditions may start at a younger age.

Participants’ responses to the open-ended questions offer some insights into understanding these results. A major reason given by participants as to why they would not want to use AUDRA was that their privacy, confidentiality, or both could be compromised in any way. This fear has been echoed in many other studies in which respondents cited data privacy concerns as reasons for not using mHealth apps [23-27]. These concerns may have trumped other factors and dampened their intention to use these apps.

Future research should further investigate the factors leading to adoption of mHealth apps, such as concerns regarding privacy. This study also has implications for practitioners. With increased efforts being made to promote the use of AUDRA, designers should first focus on making their apps convenient and easy to use. For app designers, health care professionals, and health care authorities eager to promote the adoption of AUDRA, this study suggests focusing on social influence, ensuring that the use of AUDRA is supported and encouraged by the people who matter to potential users, including their family and general practitioners along with highlighting the positive experiences of other users in their network.

Limitations

This study has some limitations to be considered when interpreting the findings. We did not have enough respondents in the follow-up survey to measure use longitudinally. In addition, respondents self-identified their AUD status, and we could not verify it; however, previous studies, through test-retest validation, have suggested overall reliability with respect to such self-identification [28] associated with AUD. Many responses also had to be discarded. The gift certificate and the ease of access associated with the internet-based survey on the AUD Facebook groups may have attracted participants who were willing to break the survey rules and may explain the high number of surveys that had to be discarded. Finally, the sample size was relatively small, which comes with associated limitations, notably in terms of statistical power.

Conclusions

This study found that performance expectancy was not significant in explaining behavioral intention to use AUDRA. Instead, social influence and effort expectancy seem to be the key factors influencing the use of such apps. As apps extend their influence into highly intimate areas of our lives, the beliefs that determine the use of technology may be shifting away from utilitarian factors such as performance. Researchers and app developers alike should keep this in mind and consider the user environment and possibly privacy concerns when developing apps.

Acknowledgments

The researchers are grateful to the kind people at the addiction treatment centers from where data were partially collected for this study. This research would not have been possible without the feedback, generosity, and participation of the directors, staff, and the clients at these centers. The researchers also extend their thanks to the remaining participants who participated on the internet and contributed a significant portion of the data used in this study.
Authors' Contributions

RM collected the data and wrote the draft. JM provided directions and developed the overall research design. PN and HB reviewed the paper.

Conflicts of Interest

None declared.

References


Abbreviations

AUD: alcohol use disorder
AUDRA: alcohol use disorder recovery apps
AVE: average variance extracted
CR: composite reliability
mHealth: mobile health
PLS-SEM: partial least squares-structural equation model
UTAUT: Unified Theory of Acceptance and Use of Technology

©Rijuta Menon, Julien Meyer, Pria Nippak, Housne Begum. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 01.04.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Evaluating User Feedback for an Artificial Intelligence–Enabled, Cognitive Behavioral Therapy–Based Mental Health App (Wysa): Qualitative Thematic Analysis

Tanya Malik1, MA; Adrian Jacques Ambrose2, MD, MPH; Chaitali Sinha1, MA

1Wysa Inc, Boston, MA, United States
2Vagelos College of Physicians and Surgeons, Columbia University, New York, NY, United States

Corresponding Author:
Tanya Malik, MA
Wysa Inc
131 Dartmouth St
Boston, MA
United States
Phone: 1 9874803442
Email: tanya@wysa.io

Abstract

Background: Digital mental health apps are rapidly becoming a common source of accessible support across the world, but their effectiveness is often influenced by limited helpfulness and engagement.

Objective: This study’s primary objective was to analyze feedback content to understand users’ experiences with engaging with a digital mental health app. As a secondary objective, an exploratory analysis captured the types of mental health app users.

Methods: This study utilized a user-led approach to understanding factors for engagement and helpfulness in digital mental health by analyzing feedback (n=7929) reported on Google Play Store about Wysa, a mental health app (1-year period). The analysis of keywords in the user feedback categorized and evaluated the reported user experience into the core domains of acceptability, usability, usefulness, and integration. The study also captured key deficits and strengths of the app and explored salient characteristics of the types of users who benefit from accessible digital mental health support.

Results: The analysis of user feedback found the app to be overwhelmingly positively reviewed (6700/7929, 84.50% 5-star rating). The themes of engaging exercises, interactive interface, and artificial intelligence (AI) conversational ability indicated the acceptability of the app, while the nonjudgmentality and ease of conversation highlighted its usability. The app’s usefulness was portrayed by themes such as improvement in mental health, convenient access, and cognitive restructuring exercises. Themes of privacy and confidentiality underscored users’ preference for the integrated aspects of the app. Further analysis revealed 4 predominant types of individuals who shared app feedback on the store.

Conclusions: Users reported therapeutic elements of a comfortable, safe, and supportive environment through using the digital mental health app. Digital mental health apps may expand mental health access to those unable to access traditional forms of mental health support and treatments.

(JMIR Hum Factors 2022;9(2):e35668) doi:10.2196/35668

KEYWORDS

digital mental health; artificial intelligence; user reviews; cognitive behavioral therapy; CBT

Introduction

The World Health Organization estimates that 450 million people worldwide have a mental disorder and a mental health gap of 1:10,000 worldwide [1]. Another report identified financial constraints and lack of serviceability as structural barriers to treatment [2]. Despite considerable progress in access to resources, the gap in mental health access, especially in industrialized countries, does not appear to have shifted [3,4]. Psychological and structural barriers to accessing mental health care, such as availability, convenience, stigma, and preference for self-care, persist and underscore the increased need for accessibility of mental health resources [5]. Digital mental health tools, such as apps and chatbots, allow for anonymity and convenience and can serve as important alternatives to bridge...
the access gap [6]. The increasing availability and usability of mobile devices may create new opportunities for overcoming the existing barriers and limited access of traditional clinical service delivery and provide customized patient-centered interventions. Similarly, smartphones and other mobile technology may have the potential to reach a greater number of users and deliver reliable and effective services, regardless of location [7,8].

For bridging the mental health access gap, understanding user experiences and attitudes toward digital mental health apps is crucial. In the context of digital mental health, the Technology Acceptance Model posits that perceived ease of use and perceived usefulness of a given technology have a positive influence on user engagement, which is required for interventions to be effective [9]. For both patients and providers, Chan et al [10] proposed criteria to use in assessing mental health apps in 4 key domains: usefulness, usability, integration, and infrastructure. In addition, acceptability of a mobile app is defined as the perceived value, usefulness, and desirability [11]. As user engagement often can be suboptimal, users’ attitudes toward the digital technology can reveal important insight into their engagement [9].

To further understand user engagement with artificial intelligence (AI)–guided digital mental health apps, this study aimed to understand user needs for impactful engagement with a digital mental health app (Wysa) by examining their user reviews. As a direct proxy for users’ attitudes toward a digital mental health app, user reviews are typically voluntary, unsolicited, and openly available on a public forum, which may provide helpful evaluations and insights into the users’ experiences and engagement. A previous qualitative analysis of user reviews on mental health apps identified design improvements, user expectations, unmet needs, and utility [12]. These user reviews are regarded as a comprehensive evaluation of the app from the user’s own perspective, which provides rich insights into the app user experience [13,14]. In addition to understanding needs for engagement, this study planned to explore the perceived value, usability, and desirability of the app as a digital mental health tool [15,16].

### Methods

#### App Background

Wysa is an AI-enabled mental health app that leverages evidence-based cognitive-behavioral therapy (CBT) techniques through its conversational interface (chatbot). The app is designed by a team based out of India, the United Kingdom, and the United States. The app is designed to provide a therapeutic virtual space for user-led conversations through AI-guided listening and support, access to self-care tools and techniques (eg, CBT-based tools), as well as one-on-one human support. The app has demonstrated efficacy in building mental resilience and promoting mental well-being through a text-based conversational interface [17]. For the time period considered (1 year), the app received an overall 4.8/5 user rating on the Google Play Store and had been downloaded by more than 2 million people. The app also exists on the Apple App Store with a similar rating of 4.9/5 but with a smaller sample of qualitative reviews. Studies have shown Wysa as having the most evidence-based treatments among other smartphone apps [18], with conversations targeting specific problems and goals [19]. The app is anonymous [20] and safe [21] and rates highly on measures of app quality [22].

#### Study Design

For direct user feedback, the authors examined reviews posted on the Google Play Store between October 2020 and October 2021, during which time, 41,114 user reviews had been received. A duration of 1 year and the use of Google Play reviews were considered to ensure a sufficiently large sample. For the analysis of descriptive feedback (n=7929), the authors codified the reasons shared by the users for their rating. User feedback in languages other than English, blanks, as well as reviews that contained 1–2-word nondescriptive statements (eg, “Really nice!”, “Awesome”, “Not interested”) were excluded (Figure 1).

The study’s primary objective was to analyze feedback content to understand the users’ experiences with engaging with a digital mental health app. As a secondary objective, the types of individuals providing feedback were also explored.
**Figure 1.** Diagram of the inclusion and exclusion criteria for user review analysis.

![Diagram of inclusion and exclusion criteria]

**Analysis**

Using a consolidated framework created by Chan et al [10], which was based on guidelines suggested by the Healthcare Information and Management Systems Society (HIMSS) and the US Federal Government for evaluating digital health apps, the written reviews were verbally grouped into the domains of the framework and further analyzed for specific themes within each [15]. To understand Wysa’s capacity to currently help and engage users, the thematic analysis examined specific domains of (1) acceptability (eg, satisfaction, matching expectations of capabilities, likelihood to recommend, and level of interactivity), (2) usability (eg, the ease, enjoyment, cultural, and demographic accessibility of use), (3) usefulness (eg, validity, reliability, effectiveness, and time required to obtain a benefit), and (4) integration (eg, security, privacy, data integration, and safety) [10,15].

Each user review was evaluated and categorized into the nonmutually exclusive domains. The domain of acceptability included statements discussing likelihood to recommend the app, frequency of use, impact of use, and reasons for use. For usefulness, mentions of what the app was being used for, specific uses (including tools and techniques), and time of use were included. Usability included mentions of ease of use, convenience, and interface features. Integration primarily consisted of reviews discussing data privacy, security, and anonymity.

The coding also enabled us to capture the emergence of the key characteristics of users who were able to receive mental health support due to increased accessibility.

**Results**

The reviews analyzed for this study were largely positive, with 6700 reviews (6700/7929, 84.50%) giving the app a 5-star rating and 2676 reviews (2676/7929, 33.75%) explicitly terming the app “helpful” or that it “helped.” Of 7929 reviews, 251 (3.17%) had a less than 3-star rating and were termed as negative reviews. The themes under the evaluation criteria aimed to capture the user experiences (Figure 2).
Acceptability

The acceptability of the app was identified through the themes emerging around engagement of exercises, interactivity of the interface, and conversational ability of AI. The users who reviewed the app rated it positively on acceptability when they found it interactive and conversational. Users reported that receiving appropriate responses to user conversations in the tools and techniques was valuable. For instance, a user compared it with other options available for self-care: “The interactive experience helped more than the journaling exercises I’ve done in the past.” Several users reported the variety of exercise-guided meditations, venting spaces, positive thinking exercises, and cognitive restructuring as important in their engagement. Additionally, the user reviews described the exercises as “educational,” “calming,” “relaxing,” and “functional.”

Users said that though “…Initially it felt silly to talk to an AI but it’s extremely well made, tailored for therapy.” Per users, the “warm, friendly, and encouraging” AI helped them recreate an environment of confiding in a friend, without having to confront the intimidation of speaking with a real person. For instance, a user mentioned “It’s really nice and I feel like I’ve been heard when others won’t listen, even if I am only talking to an AI,” and another user said it “made me feel loved and heard during a crisis.” Users also reported finding talking to the AI to be a “fun” experience, perhaps brought out by elements that keep it light and accessible by including jokes, games, bitmap images (ie, GIFs), and other interactive agents.

Users reported the interactiveness of the app as central in keeping them engaged: “The app made me laugh with its silly jokes and play.” They also found the “easy” and “instinctive” interface as a central element in a positive experience of using the app to be “easy” and “instinctive” (Figure 3). Users found it comfortable to use Wysa for numerous aspects of their well-being (Figure 4).

They also mentioned being likely to recommend Wysa to others to help with sleep, managing stress, working through anxiety, as well as to “just talk to someone.” One user said, “it listens to you and helps relieve stress and also has a lot of coping mechanisms. I definitely recommend.” Some users discussed being able to share and rely on something for “regular” support, which further contributed to the acceptability of the app. A user exemplified this by stating:

Different people may find this app useful in different ways and it doesn’t pressure you to do stuff if you aren’t ready for it (no energy or not the right type). It’s great even just as a sounding board, a place to organize your thoughts or make a to-do list, or a bit of a tiny friend in your pocket that’s not judgmental and won’t be tired of you.

However, some users did not find it helpful for their specific concerns and suggested further expansion to include these specific requirements. For instance, a user said, “Interesting concept, but it needs to learn to deal with more illnesses.”
Figure 3. Example of interactive interface.
Usability

The usability of the app was presented through emerging themes of nonjudgmentality and ease of conversation. This domain was rated positively by users as they found it to be a “safe” and “nonjudgmental” environment that is easily accessible. This feature of the app was identified from user reviews such as:

It's nice to talk to someone completely objectively. Even in therapy you feel guilty if all you do is go on and on, as is human conditioning, but being able to talk it out with Wysa is great. No judgment. Don't have to feel weird about anything.

Reviews indicated that, by conversing with AI, the pressure of performance in front of a therapist was removed, which may allow a user to express themselves more freely. Users commented on the AI interface of the “cute and approachable penguin” as helpful in cultivating a nonthreatening environment: “I love it, it's just amazing, knowing that I can talk about my inner problems to a penguin without judgment ... I love that.” In fact, 201 users commented on the “no-judgment space” as a core component in making them feel safe and comfortable.

The app usage experience was also described as “…It feels like I'm talking to a real person ... Such a friendly interface.” Users appeared to be willing to adapt their expectation in order to continue benefiting from the app, with one user saying, “a little clunky at first but once you learn how to manage it it's very helpful.”

The most common negative review of the app was for repetition and a lack of comprehension by the chatbot, which made some users feel misunderstood and sometimes want to leave the app. Language limitations felt like a barrier to others who wanted to be understood more. They expressed a want for the app in native languages, including Italian, Spanish, French, and others, with one user saying, “The application is great, but it lacks the addition of other languages ... in order to facilitate its use by all layers of society.”

Usefulness

The app's usefulness was portrayed by themes such as impact on mental health, convenience of access, and cognitive restructuring exercises. User feedback discussed that the app provided a safe and open space to challenge one's thoughts and
feelings. The usefulness of the app in this regard is captured by its efficacy in dealing with mental health concerns. A user described their experience:

I have been struggling with depression since I was a child, and was terrified of reaching out for help. Finally a few weeks ago I hit rock bottom worse than ever before. I was really scared for a while. I was seeking some form of comfort or communication but didn't want to go to anyone, not to mention money is tight. This app really helped me when I needed it most. Who knew an AI penguin would cause me to sing again?

Providing a “safe” and “anonymous” place to process one's thoughts and emotions was identified by 107 users as highly impactful.

In specific clinical utility, users reported positive effects for anxiety (n=805), stress reduction (n=480), and depressed mood (n=400). In addition, 324 users reported app usage for posttraumatic stress disorder (PTSD) symptoms, fear, and sleep issues. Users identified numerous techniques and spaces offered as being especially helpful, such as physical activity exercises, sleep stories, meditations, cognitive restructuring, and reframing exercises. Users also commented on the affordability of the app as a way to bridge mental health access: “This app really helped me especially since I don't have access to any other useful form of therapy.”

The app would seem least useful when the chatbot felt limiting or was unable to fully understand the user. Some users facing a difficult time with the app would state, “Sometimes it's frustrating that an AI can't understand you that well,” and when it couldn't understand the user’s dilemma, then it felt “empty and generic.”

Integration

The integration of the app was illustrated through the themes of privacy and confidentiality.

The app did not ask users to register themselves in any way to use the app and thus did not ask for personal details, such as demographic data. The anonymous and confidential nature of the app was a key reason for positive ratings in integration. Many users reported being satisfied with the privacy practices and finding the app “easy to trust.”

I feel really good knowing that I can talk to something completely private. I was feeling really down and I was pleasantly surprised. It was so simple yet so effective. I most definitely recommend it to someone who wants privacy and a healthy listening ear.

Characteristics of Users

The thematic analysis captured the emergence of the types of users who provided reviews in the app on Google Play Store and are also a representation of users who access digital mental health support such as Wysa. They were grouped by salient aspects of their expressed needs and concerns.

We identified 4 key groups: (1) those who self-reported having clinical issues, (2) those who reported being unable or unwilling to open up to a real person, (3) those who are financially conscious, (4) and those who are unable to access mental health professionals. Use of the app for support through self-reported diagnosis and symptoms of depression, anxiety, panic disorders, and PTSD was mentioned by 1856 individuals. They primarily used the CBT techniques and meditations on the app as a form of self-care. Another application of the app is for individuals who feel uncomfortable talking to people in their lives or who don’t have a reliable system with which to share their thoughts. They reported finding the AI-driven app useful in reducing the guilt and burden of opening up to a real person. Users also found the free nature of the app to be beneficial to reduce the burden of financial anxiety when considering mental health support. Numerous users (n=594) also reported using the app at times when they would be unable to access therapists, including when experiencing higher symptoms of depression and anxiety late at night.

Discussion

Principal Findings

This study represents one of the largest studies in understanding users’ perceptions of a digital mental health app. It looked at the acceptability, usability, usefulness, and integration of a digital mental health app, by analyzing publicly available user feedback and reviews. This approach is unique for several reasons—first, it uses user feedback that was unsolicited by the developers and promoters and is delivered in a public forum, reducing the social desirability bias that could interact in other researcher-administered evaluations. Second, the robust sample size allowed for a deeper dive of user experiences, which was previously unexplored in other studies. This approach helped to recognize the types of users of mental health apps, which helped to identify strengths and weaknesses of digital mental health tools and allowed us to better understand the gaps in services provided.

The most important findings resulting from this study are the factors that contribute to higher engagement and acceptability for a digital mental health app. Users most consistently listed the “active and available listening” element as the key to foster acceptability with the digital mental health experience. The app further cultivated the therapeutic elements via the use of an AI-based chatbot with a friendly penguin user interface. In addition, the perceived nonjudgmentality and friendliness of this interface resulted in high usability and ongoing engagement with the app.

Understanding the user experience is important to ensuring meaningful usage and clinical utility [17]. Users strongly valued the anonymity and confidentiality of the app, which are valuable strengths in any therapeutic relationship [23]. Therapeutic bonds are fostered through trust, acceptance, empathy, and genuineness and are important for their role in the effectiveness of an intervention [23] and, in a digital environment, are created by human dialogue through a conversation agent [24].

With users providing a large majority of positive reviews, the acceptability and effectiveness of Wysa as a digital mental health tool have been established [25]. Digital mental health apps can
provide important benefits, especially for supporting individuals with subclinical psychiatric symptoms [26]. The findings of this study highlighted how digital mental health apps can significantly improve the accessibility and affordability of mental health support. The characteristics of users identified helped outline those who may access and benefit from the presence of mental health apps; for example, individuals managing social anxiety symptoms of speaking face to face can find significant therapeutic value through an AI-enabled tool. In addition, mental health apps may serve as augmenting or transitioning tools during times when traditional mental health services are limited, such as after office hours, in rural settings, or in between appointments and referrals.

Limitations

Limitations to the study include the source of data, as the Apple App Store data were not considered and only the reviews on Google Play Store were addressed in this study. Further, user reviews are taken at a single point in time, and thus evidence of changes in feedback are unavailable for consideration. No demographic information was collected aside from reviews being in English. Clinical scores of users were not identified, which would otherwise have contributed to more direct understanding of the experience with the app in clinical populations. The study is also limited by lack of knowledge on the duration of app use or the rate of attrition among users due to app issues or other reasons.

Conclusions

This study utilized a user-led approach to understanding factors for engagement and helpfulness in digital mental health. User feedback was analyzed on domains of acceptability, usability, usefulness, and integration, and we found the app to be overwhelmingly positively reviewed. A key facet that emerged is the comfort and safe environment created by the nonjudgmental digital mental health tool that provides users with clinical and subclinical support. Further analysis revealed 4 predominant types of individuals who appear to be engaging in digital mental health support and who are infrequent users of face-to-face mental health services. Digital mental health apps can provide a valuable service to those unable to access mental health support. Future directions for digital mental health include improvements within the technology to cater to varied users, increasing its capacity to contribute to clinical utility.

Conflicts of Interest

TM and CS are employees of Wysa and hold equity in Wysa Inc. AJA declares no conflicts of interest related to this study.

References

7. East ML, Havard BC. Mental health mobile apps; from infusion to diffusion in the mental health social system. JMIIR Ment Health 2015 Mar 31;2(1):e10 [FREE Full text] [doi: 10.2196/mental.3954] [Medline: 26543907]


Abbreviations

AI: artificial intelligence  
CBT: cognitive behavioral therapy  
HIMSS: Healthcare Information and Management Systems Society  
PTSD: posttraumatic stress disorder

Edited by A Kashnir; submitted 13.12.21; peer-reviewed by A Hashim; comments to author 18.02.22; revised version received 24.02.22; accepted 06.03.22; published 12.04.22.

Please cite as:

Malik T, Ambrose AJ, Sinha C  
Evaluating User Feedback for an Artificial Intelligence–Enabled, Cognitive Behavioral Therapy–Based Mental Health App (Wysa): Qualitative Thematic Analysis  
JMIR Hum Factors 2022;9(2):e35668  
URL: https://humanfactors.jmir.org/2022/2/e35668  
doi:10.2196/35668  
PMID:35249886

©Tanya Malik, Adrian Jacques Ambrose, Chaitali Sinha. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 12.04.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.
Application of a Web-based Self-assessment Triage Tool During the COVID-19 Pandemic: Descriptive Study

Anna Nowicka¹²*, MD; Jakub Jaszczak²*, MD; Anna Szymanek Pasternak¹³, MD, PhD; Krzysztof Simon¹³, MD, PhD

¹Provincial Specialist Hospital them. J. Gromkowski, Wrocław, Poland
²Infermedica, Wrocław, Poland
³Department of Infectious Diseases and Hepatology, Wroclaw Medical University, Wrocław, Poland
* these authors contributed equally

Corresponding Author:
Anna Nowicka, MD
Provincial Specialist Hospital them. J. Gromkowski
Koszarowa 5
Wrocław, 51-149
Poland
Phone: 48 660427576
Email: nowickanna@gmail.com

Abstract

Background: The COVID-19 pandemic has sped up the implementation of telehealth solutions in medicine. A few symptom checkers dedicated for COVID-19 have been described, but it remains unclear whether and how they can affect patients and health systems.

Objective: This paper demonstrates our experiences with the COVID-19 risk assessment (CRA) tool. We tried to determine who the user of the web-based COVID-19 triage app is and compare this group with patients in the infectious diseases ward’s admission room to evaluate who could benefit from implementing the COVID-19 online symptom checker as a remote triage solution.

Methods: We analyzed the answers of 248,862 people interacting with an online World Health Organization–based triage tool for assessing the probability of SARS-CoV-2 infection. These users filled in an online questionnaire between April 7 and August 6, 2020. Based on the presented symptoms, risk factors, and demographics, the tool assessed whether the user’s answers were suggestive of COVID-19 and recommended appropriate action. Subsequently, we compared the sociodemographic and clinical characteristics of tool users with patients admitted to the Infectious Diseases Admission Room of J. Gromkowski Hospital in Wrocław.

Results: The CRA tool tended to be used by asymptomatic or oligosymptomatic individuals (171,226 [68.80%] of all users). Most users were young (162,432 [65.27%] were below 40 years of age) and without comorbidities. Only 77,645 (31.20%) of the self-assessment app users were suspected of COVID-19 based on their reported symptoms. On the contrary, most admission room patients were symptomatic—symptoms such as fever, cough, and dyspnea were prevalent in both COVID-19-positive and COVID-19-negative patients. COVID-19-suspected patients in the CRA tool group presented similar COVID-19 symptoms as those who presented to the admission room. These were cough (25,062/40,007 [62.64%] in the CRA tool group vs 138/232 [59.48%] in the admission room group), fever (23,123/40,007 [57.80%] in the CRA tool group vs 146/232 [62.93%] in the admission room group), and shortness of breath (15,157/40,007 [37.89%] in the CRA tool group vs 87/232 [37.50%] in the admission room group).

Conclusions: The comparison between the symptomatology of the users interacting with the CRA tool and those visiting the admission room revealed 2 major patient groups who could have benefited from the implementation of the self-assessment app in preclinical triage settings. The primary users of the CRA tool were young, oligosymptomatic individuals looking for screening for COVID-19 and reassurance early in the COVID-19 pandemic. The other group were users presenting the typical symptoms suggestive of COVID-19 at that time. The CRA tool recognized these individuals as potentially COVID-19 positive and directed them to the proper level of care. These use cases fulfil the idea of preclinical triage; however, the accuracy and influence on healthcare must be examined in the clinical setting.
KEYWORDS
COVID-19; symptom checker; preclinical triage; self-assessment tool; online applications; COVID-19 remote triage; COVID-19 self-assessment

Introduction

Background

After the outbreak of the COVID-19 pandemic, the health care systems of affected countries faced an unprecedented challenge. Ensuring the continuity of care and screening the vast number of suspected patients have put a significant strain on health care, leading to the depletion of public health resources [1,2]. Although the health system resources were transferred to provide critical services to patients suffering from COVID-19, the utilization of medical visits reduced by even 42% [2], suggesting that patients with less severe illnesses tended to avoid in-person consultation or had no possibility to attend one.

During the pandemic, especially in the early days, there was much uncertainty regarding the symptomatology and clinical course of the novel coronavirus disease. This has been reflected in the number of searches for the phrase “covid 19 symptoms” on the Google platform, which at the time of the study varied from 443,000 to 2.2 million searches per month just for the United States [3].

These uncertain times have presented an opportunity to popularize telehealth solutions in medicine. The means of remote consultations have found their way mostly in primary care as a substitute for in-person visits [4] but also as a way of remote triage of COVID-19 patients.

Triage is defined as a classification of patients according to their urgency. Remote triage uses the means of distance communication, such as telephones or interactive websites, allowing for the segregation of patients before they interact with health care professionals. Remote triage solutions have been proven helpful in telephone call centers, where they have been associated with lower in-person health care use [5]. They have also been demonstrated to be useful in the triage of COVID-19 patients, as they have reduced the number of unnecessary consultations, hence reducing the exposure of the staff to COVID-19 [6]. Web-based COVID-19 symptom checkers and triage tools have also proved useful in scheduling tests [7,8], monitoring symptoms [9-11], providing evidence-based educational value [8,9,12], and supporting self-isolation [13].

Objective

In this study, we wanted to share our findings regarding the COVID-19 risk assessment (CRA) tool. It was a World Health Organization (WHO) guidelines-based online triage tool, which assessed the risk of SARS-CoV-2 infection and returned a probable outcome with a concise recommendation of what to do next, along with evidence-based educational materials about COVID-19.

We gathered and analyzed the data of 651,757 patients interacting with the CRA tool, focusing on their demographics, risk factors, reported symptoms, possible exposure to SARS-CoV-2, and recommended triage. The aim was to establish who the main users of web-based COVID-19 symptom checkers (age, sex, comorbidities, presenting symptoms) are and who might have benefitted from implementing COVID-19 symptom checkers as preclinical triage solutions.

Since confirming the diagnosis in an online self-assessment tool was not achievable, we compared the results (sociodemographic and clinical characteristics of CRA users) with the health records of the Infectious Diseases Admission Room of J. Gromkowski Hospital in Wroclaw to establish whether and how these groups corresponded. The goal was to evaluate who could benefit from implementing this solution as preclinical triage.

Methods

Study Population

Since April 7, 2020, we have been collecting and utilizing responses from the CRA tool users. The app was developed by Infermedica company, as a non-profit project. It utilized a diagnostic algorithm designed based on WHO and Centers for Disease Control and Prevention (CDC) recommendations. The specific time frame was chosen due to periodical updates of the app questions flow. In the selected period, there were no major changes to the question flow so that the collected information could be unbiased.

Inclusion Criteria

The study population included individuals concerned about their risk of COVID-19 infection:

- Users who filled the questionnaire available through the Infermedica website between April 7 and August 6, 2020
- Users who filled the questionnaires available on third-party websites, which obtained permission to use our tool within their platforms between April 7 and August 6, 2020

Exclusion Criteria

The exclusion criteria were:

- Completing the interview in an outdated 1.0 and 2.0 version (not all providers of our tool updated their software before the beginning of the study)
- Completing the interview in a version customized for a national health system so that it was incompatible with WHO and CDC recommendations
- Not completing the whole interview
- Age below 18 years
- Completing the interview in a language other than Polish

Data Privacy and Ethical Statement

The study population consisted of 2 arms: users of the web app and patients in the admission room.
The app arm consisted of users of the web app who accepted the terms of service. All data processed through the COVID-19 risk assessment checker were anonymous and did not allow us to identify an individual based on the information provided during the interview. Informed consent to use anonymized data was provided by the users by accepting the terms of service. A privacy policy and personal data protection were applied.

The admission room arm of the study did not require ethics committee approval as a retrospective study, according to the guidelines of the local ethical compliance body [14].

COVID-19 Symptom Checker Characteristics

The CRA is a triage tool dedicated to nonprofessional users. The checkup was designed to assess whether the user's symptoms may be the result of SARS-CoV-2 infection. It had a form of a responsive web app that could be embedded within a website or an Application Program Interface (API) that can serve as a technological core for building custom apps. (An API is a set of routines, protocols, and tools for building software applications. Basically, an API specifies how software components should interact. It serves as a technological core for custom-building applications.) The flow of the interview was solely based on the official WHO guidelines for diagnosing COVID-19 [15]. The first version of the API was released on March 20, 2020 (version 1.0), followed by updates on March 25, 2020 (version 2.0), April 7, 2020 (version 3.0), and May 7, 2020 (version 4.0).

The app has been considered final from version 3.0; the set of risk factors and symptoms have reached their final form. However, the core logic of the interview, such as the flow of the interview, types of acquired data, and types of output recommendations, has been consistent from the first released version. In this study, we only considered interviews in the period between April 7 and August 6, 2020.

Medical Foundation

The CRA tool's logic was built around WHO guidelines [15] and WHO daily transmission reports [16]. The interview was designed to gather enough data to establish whether the user falls into any of the 3 categories mentioned in said guidelines as “Suspected case” for COVID-19; therefore, the reported symptoms may have resulted from SARS-CoV-2 infection.

For this reason, the interview consisted of 3 sets of questions that could be grouped into 3 categories:

- Risk factors and symptoms
- Places of residence and travel
- Contact with possible COVID-19 cases

In some cases, when this information was unnecessary to make a diagnosis, some questions were omitted.

Data Analysis

The majority of the data were compared and presented with the use of descriptive statistics. Inferential statistics had to be omitted because of the significant differences in both compared populations and vastly different sample sizes. We decided to only use statistical analysis to compare comorbidities related to COVID-19 in both CRA and admission room groups. In CRA, P values were calculated with the test of proportions and in the admission room, with the Fisher exact test.

Screen Deep Dive

The interview consisted of up to 8 consecutive screens. Not every screen had to be included; this is the maximum number of screens that the user could have been exposed to. If the patient reported emergency evidence (ie, acute dyspnea), the interview was terminated with an instruction to call an ambulance. The screens in the display order were “Welcome & Terms of Service,” “Age and Sex Selection,” “Risk Factors,” “Symptoms,” “Red Flags,” “Possible Exposure to COVID-19,” “Travel and Residency,” and “Outcome.”

Nine risk factors were included to inquire about the user’s chronic illnesses and overall medical condition: diseases or drugs that weaken the immune system, obesity, long-term stay at a care facility or nursing home, diabetes, cancer, cardiovascular disease, history of chronic lung disease, history of chronic liver disease, and history of chronic kidney disease. Some of these comorbidities have been described as negatively impacting COVID-19 infection outcomes [17]. We also included risk factors described in the Pneumonia Severity Index (PSI) as a negative prognostic factor indicating the need for hospitalization [18].

The symptom screens were oriented on inquiring about users’ symptoms that should raise clinical suspicion for COVID-19 according to WHO guidelines [15]. There was a list of 11 symptoms users could choose from: fever, cough, shortness of breath, fatigue, muscle pain, chills, headache, diarrhea, nausea, sore throat, and impaired taste or smell. Furthermore, the interview focused on assessing red flags—immediate health threats to the user that should yield in cessation of the interview. To do so, the user was asked about rapid symptom deterioration, tachypnea, or hemoptysis.

There were 6 possible outcomes of the interview, which referred to the possibility of COVID-19 infection and the severity of symptoms:

- COVID-19 suspected, serious: “Call the emergency number. Avoid all contact.”
- COVID-19 suspected, nonserious: “Consult your health care provider. Avoid all contact.”
- Contact with COVID-19, no symptoms: “Quarantine.”
- Non-COVID-19, serious: “Call a doctor.”
- Non-COVID-19, nonserious: “Stay home and monitor your symptoms.”
- Asymptomatic: “Follow preventive measures.”

The extensive screen description and decision tree logic can be browsed in Multimedia Appendix 1.

Comparison Group: Admission Room Analysis

To compare individuals completing the survey with real patients diagnosed with COVID-19 by health care professionals, we turned to the Infectious Diseases Admission Room of J. Gromkowski Hospital in Wroclaw. We analyzed 291 cases of patients visiting the admission room between April 7 and August 6, 2020. All the patients reporting to the admission room were
suspected of COVID-19 infection; no other cases of infectious diseases were consulted in the admission room at that time. They may have been brought to the admission room by ambulance, referred by the primary care physician, or admitted by themselves. We excluded patients below 18 years of age.

Each patient was interviewed and examined by the physician working in the admission room. The interview consisted of fixed elements, such as current symptoms, comorbidities, medication, history of travel, contact with COVID-19-positive persons, and workplace and family interview. Blood analysis, chest X-rays, and COVID-19 swabs were obtained in most cases.

The patient's history and examination, along with the additional tests, allowed them to decide on admission to the hospital or discharge. After 24 hours, the results of the COVID-19 genetic test (reverse transcription polymerase chain reaction [RT-PCR] from nasopharyngeal or pharyngeal swabs) were available, which allowed reaching the final diagnosis.

Setting

J. Gromkowski Hospital in Wrocław, Lower Silesian Voivodeship, Poland, is 1 of the specialist hospitals in that city. There are 2 infectious disease wards in the hospital. The Infectious Diseases Admission Room serves as the place for preliminary triage, diagnosis, and treatment of incoming patients suspected of contracting infectious diseases. During the COVID-19 pandemic, it served as the main consultation facility of COVID-19-suspected cases.

Population

In this study, we analyzed the Infectious Diseases Admission Room cases between April 7 and August 6, 2020. We focused on the set of reported symptoms, comorbidities, contact with COVID-19 cases, and travel history. Our goal was to determine the patient profile, meaning assessing the set of symptoms connected with COVID-19 cases compared to non-COVID-19 cases.

Finally, we wanted to compare the sociodemographic and clinical characteristics of hospital patients and the ones completing the self-assessment interview.

Symptoms

In the study, we screened for 8 symptoms that are suggestive of COVID-19 infection: cough, fever, dyspnea, diarrhea, myalgia, rhinorrhea, taste and smell abnormalities, and pharyngeal pain.

Results

Demographics and Groups Characteristics

CRA Tool

Of the 697,903 individual interviews performed on the CRA tool between April 7 and August 6, 2020, a total of 248,862 (35.66%) individual interviews met the inclusion criteria. Most of these interviews came from the government portal of the Polish Ministry of Health, which embedded the app within its website [19]: 117,311 (47.14%) of all interviews. In addition, 91,805 (36.89%) interviews were performed on the original CRA website [20], and 17,767 (7.14%) interviews were performed on the COVID-19 mobile app commissioned by the Polish Ministry of Health. Other notable institutions adopting the CRA tool and providing us interviews analyzed in the study included PZU Zdrowie (Polish biggest private health care provider), Dovera (private health care provider in Slovakia), Global Excel (medical assistance company operating in the U.S. and Canada), and others [21]. The CRA tool is offered in 37 languages in total: Polish, English, Slovak, Ukrainian, Portuguese-Brazilian, and Russian are the most popular languages. However, only Polish-speaking users met the inclusion criteria (Figure 1).

Most of the respondents were between 18 and 40 years old (n=158,998 [63.89%] of all respondents). The least prevalent were users between 80 and 90 years old (n=498, 0.2%). The mean age was 37 years. The study included 130,966 (52.63%) males and 117,896 (47.37%) females (Figure 2).
Figure 1. Finished CRA interviews daily (blue line); for comparison, daily number of new diagnosed COVID-19 cases in Poland (red line). CRA: COVID-19 risk assessment.

![Graph showing finished CRA interviews and new COVID-19 cases in Poland]

Figure 2. Age and sex distribution of admission room patients (N=291).

![Bar chart showing age distribution and pie chart showing sex distribution]

**Admission Room**

The study included 291 patients who visited the Infectious Diseases Admission Room of J. Gromkowski Hospital in Wroclaw between April 7 and August 6, 2020. There were 152 (52.23%) women and 139 (47.77%) men enrolled in the study. Most of the patients were between 41 and 70 years old. The mean age was 58 years; the median age was 60 years (Figure 3).
Outcomes and Triage Results

CRA Tool

Among the users of the CRA tool, the most common interview result was “asymptomatic” or “Follow preventive measures,” which was displayed to 98,081 (39.41%) of the 248,862 users. This subgroup consisted of users who answered the questionnaire but denied having any symptoms or COVID-19 exposure.

The second-most common triage outcome was “non-COVID-19, nonserious” or “Stay home and monitor your symptoms” for 73,145 (29.39%) of the 248,862 users. This subgroup comprised users who answered the questionnaire and reported only mild symptoms, such as fatigue, muscle pain, chills, headache, diarrhea, nausea, sore throat, and impaired taste or smell, but denied having any COVID-19 exposure (contact or travel). These users were not suspected of SARS-CoV-2 infection according to the diagnosing rules proposed by WHO at that time [15].

Both these groups added up to 171,226 (68.80%), which made them the majority of the CRA tool users. See Tables 1 and 2 for details of the CRA tool group.

The third-most common triage outcome was “Call the emergency number,” which was recommended to 30,494 (12.25%) of the users. These were referred to as “COVID-19 suspected, serious” cases. Users who received that recommendation reported red-flag symptoms indicating respiratory distress or potentially severe infection (shortness of breath in the elderly, tachypnea, hemoptysis, high-grade fever, rapid symptom deterioration) and confirmed potential COVID-19 exposure.

Of the 248,862 users, 21,980 (8.83%) were classified as “Non-COVID-19, serious”: they received a “Call a doctor” recommendation. These users were not suspected of SAR-CoV-2 infection, because they had not met the WHO criteria of the suspected case at the time [15], but were advised to obtain a teleconsultation due to potentially severe symptoms: shortness of breath, high-grade fever, and fever and cough in the elderly.

The least prevalent group was the “COVID-19 suspected, nonserious” or “Consult your health care provider. Avoid all contact,” displayed to 9513 (3.82%) users. This group reported symptoms and COVID-19 exposure suggestive of SARS-CoV-2 infection but denied having potentially serious symptoms calling for an in-person consultation. They were advised to self-isolate and undergo a COVID-19 test.
Table 1. Distribution of CRA interview outcomes (N=248,862).

<table>
<thead>
<tr>
<th>Triage</th>
<th>Patients, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptomatic</td>
<td>98,081 (39.41)</td>
</tr>
<tr>
<td>Non-COVID-19, nonserious</td>
<td>73,145 (29.39)</td>
</tr>
<tr>
<td>COVID-19 suspected, serious</td>
<td>30,494 (12.25)</td>
</tr>
<tr>
<td>Non-COVID-19, serious</td>
<td>21,980 (8.83)</td>
</tr>
<tr>
<td>Quarantine</td>
<td>15,649 (6.29)</td>
</tr>
<tr>
<td>COVID-19 suspected, nonserious</td>
<td>9513 (3.82)</td>
</tr>
</tbody>
</table>


Table 2. Distribution of CRA interview outcomes: matrix of the clinical suspicion of COVID-19 (N=248,862).

<table>
<thead>
<tr>
<th>Severity of presented symptoms</th>
<th>COVID-19 suspected, n (%)</th>
<th>Non-COVID-19, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serious</td>
<td>30,494 (12.25)</td>
<td>21,980 (8.83)</td>
</tr>
<tr>
<td>Nonserious</td>
<td>9513 (3.82)</td>
<td>171,226 (68.80)</td>
</tr>
</tbody>
</table>


Admission Room

Of the 291 patients, 232 (79.73%) tested positive for COVID-19 and 59 (20.27%) tested negative for COVID-19. Of the 152 women, 126 (82.89%) were COVID-19 positive and 26 (17.11%) were COVID-19 negative. Of the 139 men, 106 (76.26%) were COVID-19 positive and 33 (23.74%) were COVID-19 negative.

Most of the patients (n=167, 57.39%) of the admission room group were classified by consulting physicians as patients in good general condition, 85 (29.21%) of the patients were judged to be in moderate general condition, 30 (10.31%) were in a bad general condition, and 9 (3.09%) were in a severely bad general condition.

Comorbidities

The number of reported comorbidities in the self-assessment app was 71,515; at least 1 risk factor was reported in 71,523 (28.74%) of the interviews. In other words, in 177,339 (71.26%) of the interviews, users did not report any comorbidity.

The most frequently reported comorbidity in both the CRA tool users and the admission room patients was cardiovascular disease, defined as hypertension, coronary disease, or heart insufficiency and confirmed by 37,628 (15.12%) of 248,862 CRA tool users and 138 (47.42%) of 291 admission room patients.

The distribution of other comorbidities shaped quite similarly between the 2 compared groups:

- In the CRA tool group, the other common risk factors were chronic lung disease (8337/248,862, 3.35%) and diabetes (5998/248,862, 2.41%).
- In the admission room group, the other common risk factors were diabetes (56/291, 19.24%), cancer (active neoplasms of all types, including of hematological origin; 30/291, 10.31%), and chronic lung disease (22/291, 7.56%).

A relatively high percentage of people reporting immunosuppression in the CRA tool group (weakened immune system; 14,708/248,862 [5.91%] of users) compared to the admission room group (6/291, 2.06%) suggests this risk factor might have been misinterpreted and misused despite the extensive description explaining the nature and examples of immunosuppression (available in Multimedia Appendix 2).

In general, admission room patients more often were burdened with comorbidities compared to CRA tool users. This can be explained by a higher average age of admission room patients compared to CRA tool users (Table 3).
Table 3. Distribution of comorbidities\textsuperscript{a} in the CRA\textsuperscript{b} tool and admission room groups.

<table>
<thead>
<tr>
<th>Comorbidities</th>
<th>CRA tool</th>
<th>Admission room</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>COVID-19 positive (N=193,206), n (%)</td>
<td>COVID-19 negative (N=40,007), n (%)</td>
</tr>
<tr>
<td>Cardiovascular diseases</td>
<td>9346 (23.36)</td>
<td>26,296 (13.61)</td>
</tr>
<tr>
<td>Diabetes</td>
<td>1680 (4.20)</td>
<td>4012 (2.08)</td>
</tr>
<tr>
<td>Current cancer</td>
<td>818 (2.04)</td>
<td>1517 (0.79)</td>
</tr>
<tr>
<td>Diagnosed chronic lung disease</td>
<td>2461 (6.15)</td>
<td>5425 (2.81)</td>
</tr>
<tr>
<td>History of chronic liver disease</td>
<td>1064 (2.66)</td>
<td>2140 (1.11)</td>
</tr>
<tr>
<td>History of chronic kidney disease</td>
<td>967 (2.42)</td>
<td>1851 (0.96)</td>
</tr>
<tr>
<td>Weakened immune system</td>
<td>4309 (10.77)</td>
<td>9629 (4.98)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Overall comorbidities: There were 20,645 comorbidities in COVID-19 positives and 50,870 comorbidities in COVID-19 negatives in the CRA tool group. There were 243 comorbidities in COVID-19 positives and 24 comorbidities in COVID-19 negatives in the admission room group.

\textsuperscript{b}CRA: COVID-19 risk assessment.

\textsuperscript{c}P values for CRA: test of proportions.

\textsuperscript{d}P values for the admission room: Fisher exact test.

**Symptom Distribution**

Overall, the most commonly reported symptoms differed between the CRA tool and the admission room groups. CRA interviews were dominated by mild symptoms, such as fatigue (61,544/248,862, 24.73%), cough (54,575/248,862, 21.93%), and headache (45,417/248,862, 18.25%). Meanwhile, the admission room patients presented with more serious symptoms, such as fever (175/291, 60.14%), cough (168/291, 57.73%), shortness of breath (114/291, 39.18%), and fatigue and muscle pain (59/291, 20.27% for both).

In the admission room group, the distribution of the most common symptoms among COVID-19-positive (232/291, 79.73%) and COVID-19-negative (59/291, 20.27%) patients was fairly similar: fever (n=146 [62.9%] of COVID-19 positives, n=29 [49.2%] of COVID-19 negatives), cough (n=138 [59.5%] of COVID-19 positives, n=30 [50.8%] of COVID-19 negatives), and shortness of breath (n=87 [37.5%] of COVID-19 positives, n=27 [45.8%] of COVID-19 negatives).

In contrast, the presentation of the COVID-19-suspected and COVID-19-nonsuspected individuals differed substantially. COVID-19-suspected users commonly reported symptoms such as fever, cough, and shortness of breath, while COVID-19-nonsuspected users commonly reported headache, cough, and fatigue. For details see Table 4.
# Table 4. Symptom and risk factor distribution of CRA⁴ tool users and admission room patients.

<table>
<thead>
<tr>
<th>Symptom or risk factor</th>
<th>CRA tool</th>
<th>Admission room</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>COVID-19 positive (N=40,007), n (%)</td>
<td>COVID-19 negative (N=193,206), n (%)</td>
</tr>
<tr>
<td>Cough</td>
<td>25,062 (62.64)</td>
<td>29,521 (15.28)</td>
</tr>
<tr>
<td>Fever</td>
<td>23,123 (57.80)</td>
<td>20,292 (10.50)</td>
</tr>
<tr>
<td>Symptoms getting worse quickly</td>
<td>19,816 (49.53)</td>
<td>0</td>
</tr>
<tr>
<td>Shortness of breath</td>
<td>15,157 (37.89)</td>
<td>12,717 (6.58)</td>
</tr>
<tr>
<td>Faster breathing</td>
<td>12,964 (32.40)</td>
<td>0</td>
</tr>
<tr>
<td>Fatigue</td>
<td>5987 (14.96)</td>
<td>52,630 (27.24)</td>
</tr>
<tr>
<td>Headache</td>
<td>4497 (11.24)</td>
<td>38,115 (19.73)</td>
</tr>
<tr>
<td>Sore throat</td>
<td>3975 (9.94)</td>
<td>35,645 (18.45)</td>
</tr>
<tr>
<td>Muscle pain</td>
<td>3351 (8.38)</td>
<td>27,015 (13.98)</td>
</tr>
<tr>
<td>Coughing up blood</td>
<td>2006 (5.01)</td>
<td>0</td>
</tr>
<tr>
<td>Chills</td>
<td>1906 (4.76)</td>
<td>13,740 (7.11)</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>1242 (3.10)</td>
<td>14,109 (7.30)</td>
</tr>
<tr>
<td>Contact with infected person</td>
<td>1005 (2.51)</td>
<td>0</td>
</tr>
<tr>
<td>Nasal catarrh</td>
<td>954 (2.38)</td>
<td>6134 (3.17)</td>
</tr>
<tr>
<td>Loss of smell or taste</td>
<td>947 (2.37)</td>
<td>6034 (3.12)</td>
</tr>
<tr>
<td>Nausea</td>
<td>911 (2.28)</td>
<td>10,599 (5.49)</td>
</tr>
<tr>
<td>No contact with infected person</td>
<td>0</td>
<td>193,206 (100)</td>
</tr>
</tbody>
</table>


Comparative Results

Fever and cough were the most commonly reported symptoms of COVID-19 in CRA tool users and admission room patients: fever occurred in 23,123/40,007 (57.80%) and 146/232 (62.93%) of the studied groups, respectively, while cough occurred in 25,062/40,007 (62.64%) and 138/232 (59.48%) of the studied groups, respectively. Pneumonia, characterized as the presence of fever, cough, and dyspnea, has been proven to be the most prevalent clinical presentation of COVID-19 in many studies [22-25].

Cardiovascular disease and diabetes occurred significantly more commonly in the COVID-19-positive than in the COVID-19-negative group both in the CRA tool (9346/40,007 [23.36%] vs 26,296/193,206 [13.61%] for cardiovascular disease, P<.001; 1680/40,007 [4.20%] vs 4012/40,007 [2.08%], P<.001 for diabetes) and in the admission room (125/232 [53.88%] vs 13/59 [22.03%], P=0.047) for cardiovascular disease; 51/232 [21.98%] vs 5/59 [8.47%], P<.001 for diabetes) group.

Anosmia or ageusia (2/59, 3.39%) occurred more frequently in the admission room group in COVID-19-positive than in COVID-19-negative patients. In the app, we did not observe a similar finding, probably due to the rapid cessation of the interview in high-triage scenarios.

Anosmia or ageusia occurred more frequently in mild than in severe COVID-19 in the CRA tool group (3849/40,007 [9.62%] vs 40/40,007 [0.10%]). This is consistent with studies suggesting that olfactory and gustatory disturbances are among the most commonly reported symptoms in mild-to-moderate COVID-19 [26].

The average age of users of the COVID-19 self-assessment app was 37 years, whereas the average age of admission room patients was 58 years.

Fatigue, chills, nausea, and sore throat did not turn out to be diagnostically relevant for diagnosing COVID-19. In both CRA tool and admission room groups, they occurred more frequently in non-COVID-19 individuals.

Discussion

Principal Findings

The CRA tool ceased to be supported on August 16, 2021. As of now, most of the COVID-19 diagnostics are run by the Infermedica artificial intelligence (AI) engine [27], and the CRA tool is supported only in selected use cases (ie, the Polish Ministry of Health) [19].

The CRA tool, as it served as a means of screening and self-education, did not substitute for consultations in the admission room for symptomatic users. The tool could not confirm or exclude SARS-CoV-2 infection, as it cannot perform a laboratory examination. Hence, it does not substitute for physicians’ interactions. However, our tool exercised the purpose...
of remote triage. CRA did not overlook truly symptomatic cases; users with potentially worrisome symptoms, such as fever or shortness of breath, were identified and advised to obtain a consultation or schedule a COVID-19 test.

The compared groups—one that completed the online interview and one that reported to the hospital—differed in age distribution, the presence of risk factors, and probably the severity of symptoms reported. The difference between both groups impacted the results of the study, but it also showed some limitations of remote diagnostic tools, such as CRA—as patients potentially the most vulnerable to COVID-19 are also the least prevalent group accessing the internet for a health checkup. It is observed, however, that younger patients also suffer from COVID-19 infection, and with the next waves of pandemics, infections in young adults will become more prevalent [28]. This growing group of patients could have benefited from remote triage assessment tools, such as CRA.

Taste and smell disorders occurred more commonly in the admission room group than in the CRA tool group (39/232 [16.81%] vs 947/40,007 [2.37%] for COVID-19-suspected individuals). In search of a possible explanation of this finding, we turned to the logic of WHO guidelines used in the CRA tool at that time. They did not distinguish smell and taste disorders as key diagnostic factors [15]. Once the importance of symptoms such as smell and taste disorders came to the attention of academics [29], WHO reflected these findings in the updated guidelines for suspecting COVID-19 infection (on August 7, 2020). WHO emphasized adjacent symptoms, such as diminished taste or smell, and reduced the significance of fever in suspecting COVID-19 infection. The newer versions of the CRA tool, not described in this paper, follow the guidelines, increasing their diagnostic importance.

It was not possible to assess the actual number of false-negative cases in the CRA tool due to a lack of data. However, we know that among the admission room records, 31 (13.78%) of 225 patients did not present with fever or dyspnea but still tested positive for COVID-19. These patients would have been classified as non-COVID-19 cases by the app.

Concomitant symptoms, such as fatigue, headache, and diarrhea, occurred infrequently in severe COVID-19-positive cases in the app. This may have been caused by the premature cessation of the interview for safety reasons.

The overall number of COVID-19-suspected cases in the CRA tool was 40,007 (16.08%) of 248,862 individual interviews. This number corresponds with the number of scheduled tests for novel coronavirus because in both these cases, we deal with the suspicion of COVID-19 based on presented history and symptoms. During a similar period, between May 11 and August 3, 2020, there were 17,864,205 tests for SARS-CoV-2 performed [30].

Limitations of the Study

Possible Misinterpretation of Red-Flag Questions

The outcomes of the self-assessment triage tool highlighted room for improvement with regard to phrasing questions in web apps for the common user. The “symptoms getting worse quickly” red flag was meant to pinpoint a swiftly deteriorating user’s general condition, which is a premise for hospitalization. However, a comparable number of confirmative and declined answers suggest that many of these answers could have been false positives. This answer might have been overly reported by the respondents, who may have misinterpreted its scope. In many cases, this occurrence may have led to the overtriage of urgent COVID-19 case recommendation (“Call the emergency number”).

Bias of the Sample

As the tool was publicly available to everyone and no check-in or login was required, there is a possibility that some users did not present the symptoms they reported and used the tool only out of curiosity or for educational purposes. However, this bias is probably limited by the size of the group tested with the self-assessment tool.

More Detailed Screening in the Admission Room Sample

Screening in the admission room is always more exhaustive than in any self-assessment tool. There are a couple of contributing factors:

- Physical examinations cannot be substituted by any questions asked by the symptom checker.
- A general appearance provides valuable clinical information to experienced clinicians.
- There is a closed set of symptoms to choose from in the CRA tool.
- After detecting a potential red flag, the tool is designed to terminate the interview without inquiring about concomitant symptoms.

Conclusions

Comparing the symptomatology of users interacting with the CRA tool and those visiting the admission room revealed 2 major patient groups that could have benefited from implementing the self-assessment app in preclinical triage settings.

The first group were patients with typical COVID-19 symptoms: cough and fever, sometimes accompanied by shortness of breath, tachypnea, fatigue, headache, and muscle pain. Some of these patients had additional comorbidities, such as diabetes or cardiovascular disease, that could have impacted the clinical course of COVID-19 [17]. The CRA tool could recognize patients with such symptoms as potentially COVID-19 positive and directed them to the proper care. The CRA tool was accurate in identifying patients at risk: every patient reporting a potential red flag was meant to pinpoint a swiftly deteriorating patient’s general condition, which is a premise for hospitalization. However, a comparable number of confirmative and declined answers suggest that many of these answers could have been false positives. This answer might have been overly reported by the respondents, who may have misinterpreted its scope. In many cases, this occurrence may have led to the overtriage of urgent COVID-19 case recommendation (“Call the emergency number”).

The other group were patients with no symptoms suggesting COVID-19 infection but still searching for answers as to whether they could be infected and what they should do. Oligosymptomatic and asymptomatic users, who constituted the majority of individuals interacting with the tool, were educated about their symptoms and advised to refer to the primary care in the case of symptom worsening. CRA has played
an educational role in advising on isolation precautions, organizing quarantine, and referring for further reading using evidence-based sources, such as WHO and the CDC.

It seems that these types of solutions may serve as health information hubs for oligosymptomatic individuals and means of remote triage for a vast audience. They possess the ability to identify patients at risk, providing them with next-step recommendations, as well as sifting out asymptomatic individuals, providing them with evidence-based education materials. Such patients were the most prevalent (171,226 [68.80%] of the 248,862 CRA tool users).

As the study did not examine the intention of the user, it is uncertain what portion of such patients would visit a health care professional unnecessarily; further studies are required to assess the exact impact of online tools on reducing unnecessary visits. Still, as we observed oligosymptomatic patients visiting the hospital admission room, it can be assumed that some portion of such visits could be prevented by providing reassuring information to the patient through the online tool.

Acknowledgments
The Department of Infectious Diseases and Hepatology of the Wroclaw Medical University provided funding for the statistical analysis. Infermedica provided funding for the submission fee.

Authors' Contributions
AN and JJ conceived and presented the idea. JJ collected and analyzed the data regarding the self-assessment tool. AN collected and analyzed the data regarding the Infectious Diseases Admission Room. AN and JJ wrote the first version of the manuscript. ASP and KS supervised the work, provided new ideas regarding Discussion and Conclusion sections, and helped with result interpretation. All authors discussed the results and contributed to the final manuscript.

Conflicts of Interest
AN and JJ are affiliated professionally with Infermedica as medical consultants. They have contributed to creating the COVID-19 risk assessment (CRA) tool by outlining and adapting the medical foundation based on World Health Organization (WHO) guidelines for COVID-19 surveillance. They did not receive any compensation for the study.
AN and ASP are physicians at the First Infectious Diseases Department in J. Gromkowski Hospital in Wroclaw. ASP is an assistant at the Department of Infectious Diseases and Hepatology of the Wroclaw Medical University.
KS is physician and the head of the First Department of Infectious Diseases in J. Gromkowski Hospital in Wroclaw, as well as the head of the Department of Infectious Diseases and Hepatology of the Wroclaw Medical University.

Multimedia Appendix 1
Extensive screen description and decision tree logic.

[DOCX File, 1795 KB - humanfactors_v9i2e34134_app1.docx ]

Multimedia Appendix 2
Distribution of symptoms and comorbidities in the CRA tool, displayed by the triage outcome. CRA: COVID-19 risk assessment.

[XLSX File (Microsoft Excel File), 13 KB - humanfactors_v9i2e34134_app2.xlsx ]

References
3. All-in-One SEO Toolset. URL: https://ahrefs.com/ [accessed 2022-03-17]


Abbreviations

API: Application Program Interface
CDC: Centers for Disease Control and Prevention
CRA: COVID-19 risk assessment
WHO: World Health Organization

©Anna Nowicka, Jakub Jaszczak, Anna Szymanek Pasternak, Krzysztof Simon. Originally published in JMIR Human Factors (https://humanfactors.jmir.org), 04.04.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Human Factors, is properly cited. The complete bibliographic information, a link to the original publication on https://humanfactors.jmir.org, as well as this copyright and license information must be included.